Essays in Empirical Microeconomics

Pascal Achard

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 13 December 2019

European University Institute
Department of Economics

Essays in Empirical Microeconomics

Pascal Achard

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Andrea Ichino, EUI, Supervisor
Prof. Michèle Belot, EUI
Prof. Yann Algan, Sciences Po
Prof. Eleonora Patacchini, Cornell University

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, Pascal Achard certify that I am the author of the work "Essays in empirical microeconomics" I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

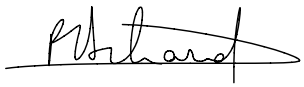
I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Statement of inclusion of previous work (if applicable):

I confirm that chapter 2 was jointly co-authored with Ms. Eva Johansen and I contributed 50% of the work.

Signature and Date:



05.09.2019

Acknowledgements

I would like to thank my supervisors Andrea Ichino and Michèle Bélot for their guidance throughout the PhD. I am particularly indebted to Michèle for her help in getting access to the data on which (most of) this thesis is based. I am also thankful to Jérôme Adda for the many discussions we had when I was at the EUI.

I would like to express my gratitude to Yann Algan and Eleonora Pattachini for being part of my jury. I benefited from many fruitful discussions with Fabian Lange and Joseph Doyle when they were visiting the department.

A special thank goes to Inés Berniell who always took the time to discuss and comment my (very preliminary) research ideas.

I would like to thank Sigrid Suetens for having given me the opportunity to work as a post-doc directly after the fourth year in Florence. This allowed me to smoothly transition out of the PhD, i.e. not to go too early on the job market! Madina, I will always be grateful that you forwarded me the job add!

During my time at the EUI, I had the privilege to meet many nice colleagues, many of whom became dear friends. Among others, Johannes, Sergey, Agnès, Chiara, Karol, Chima, David, Melanie, Jeanne ... and many more. Thanks for the challenging research discussions as well as the fun moments, but above all thanks for sharing the "palle" of the PhD.

I particularly enjoyed serving as a rep for one year and hope that it was to the benefits of other students. I wish to thank my (former) co-reps, Johanna, Matteo and Mathjis for the work we all put in.

I would like to thank the EUI presidential fund for its financial support.

Cette thèse est dédiée à mon frère Frédéric.

Abstract

This thesis is a collection of independent essays in applied microeconomics.

In the first chapter, I investigate if growing up in an ethnic enclave slow down the adoption of natives cultural behaviour by immigrants. To measure cultural behaviour, I use administrative data on usage of contraceptives by women aged 15 to 20. To observe exogenous variation in the ethnic concentration of (close) neighbourhoods, I rely on the random allocation of asylum seekers to government run accommodation in the Netherlands during the period 1996 to 2012. Although behaviours do converge with time, neighbourhood ethnic composition has no effect on this process.

In the second chapter, co-authored with Eva Johansen, we study if teenagers decision to use contraceptives is influenced by peers. To identify peer effects, we rely on cross-cohort variation in students usage in Danish high-schools. To address the reflection problem, we focus on the influence of older cohorts on younger ones. Contraception not being prevalent among young women with a non-Western background, its usage is a good measure of cultural adaptation. Looking at the effect of different peers group is indicative of which is influential. Immigrant teenagers adapt their behaviours to what other immigrants (but not what other natives) do. Their probability of using contraceptives and of having an abortion becomes lower, but not their likelihood of being treated for chlamydia.

In the third chapter, I study the influence of pre-migration social background on the economic assimilation of immigrants. I use unique French survey data to trace family histories over three generations, both in the sending country before migration and later in France. Pre-migration socio-economic status is key in explaining the educational achievements of second generation immigrants. Holding the origin country fixed, it is as important as father's occupation in the destination country. After an initial loss at migration, the first generation regains human capital more slowly than the second generation develops its own. In a simple model of human capital accumulation, this can be due to (i) parents investing more in their children than in themselves or (ii) the productivity of the two investments being different. The latter channel is supported empirically.

Contents

Chapter 1

Does growing up in an ethnic enclave slow down the adoption of natives cultural behaviour?

1 Introduction

In 2014, the European Social Survey added a module to understand which characteristics natives considered important to accept immigrants[1]. The most important factor was not race (mentioned as very important by 3.3% of respondents), neither was it religion (mentioned by 6.4%), nor education level or skills needed in the country (respectively 11.7% and 15.6% of respondents). It was the willingness to commit to the way of life of natives (mentioned by 24.6% of respondents).

Recent research Tabellini (forthcoming) has found that at the time of mass migration in the US, natives could react negatively towards immigrants, by voting for more anti-immigration parties, although their arrival increase employment and wages. This political backlash was stronger when new immigrants were culturally different, in terms of origin countries, from the population already established.

These two sets of evidence, from different continents and different periods, show that cultural distance with natives is a lasting political and social concern. It illustrates how important it is to know if, with time, immigrants adopt the mainstream culture of their new country. The existence of ethnically concentrated neighbourhoods is both a symptom of limited ethnic mixing and potentially a factor slowing down the cultural convergence process.

It is difficult to provide scientific evidence on the adoption of cultural behaviour and how neighbourhood affects it for two reasons: How to measure culture? Especially how to isolate a

[1] Answer to the question "Please tell me how important you think each of these things should be in deciding whether someone born, brought up and living outside [country of interview] should be able to come and live here", a series of characteristics are given and respondents have to say if this characteristic is not important at all (modality 0) up to extremely important (modality 10).

dimension in which immigrants and natives differ? How to make inference on the effects of ethnic concentration considering that residential choices are highly endogenous?

In this paper, I tackle these two problems. To measure cultural behaviour, I use administrative data on usage of contraceptives by women aged 15 to 20 years old. Since taking contraceptives relates to female sexuality and control over fertility, its usage differ widely between origin countries. To provide exogenous variation in neighbourhood characteristics, I focus on asylum seekers who lived in dedicated public housing when they arrived in the Netherlands. Since families could not choose where they would live and were allocated in a quasi random fashion over the country, this assignment provides an exogenous variation in neighbourhood characteristics. The combination of the outcome and the identification strategy is ideal to study immigrants cultural behavior and ethnic enclaves.

In the context of immigration from non-Western countries coming to Europe, cultural distance can be measured in different ways: whether immigrants give native sounding names to their children, Algan et al. (2013), their attitudes towards mixed marriages, Bisin et al. (2008), or pre marital female sexuality, its occurrence and its acceptance by the family and by young women themselves[2]. The sexual revolution of the 1960s-1970s, which happened in the Western World, and changed views on the sexual behaviour of young women, did not reach most part of the World. In that sense, it is an element specific to European/Western culture that young immigrants from a non-Western background may adopt or not.

Obviously, there is no administrative data on pre-marital sex. I can rely, however, on information about usage of contraceptives. It does not match one to one information on sexual behaviour, for instance people can buy condoms in a pharmacy, have pre-marital sex and not use contraceptives. However, the correlation between being sexually active and using contraceptives can only be positive (at least weakly). In itself using contraceptives is a way for women to take control over their fertility. It is an empowerment device Goldin and Katz (2002); Bailey (2006) and its wide usage by teenagers capture well the evolution which happened in the Western World over the last decades. In the data, there are large differences in usage between natives and non-Western immigrants. By age 17, 65% of natives have used contraceptives at least once while only 26% on non-Western immigrants did. At age 20, these numbers are 87% and 50%.

It is important to emphasize that this paper is about behaviour and not about identity. I look at what immigrant women do (or do not) rather than how much they are attached to their religion

[2]In a descriptive study on second generation immigrants in Canada, Lalonde and Giguère (2008) singles out attitudes towards pre-marital sex as one where immigrants from non Western origins feel torn between their cultural heritage and the mainstream views in the country. "Of greater interest, however, were ratings about the perceived appropriateness of engaging in premarital sex if involved in a loving relationship. As expected, South Asian and Chinese Canadians perceived premarital sexual intercourse as less appropriate than did their Canadian peers, although this difference was more marked for South Asian Canadians. More importantly, the ratings of South Asian Canadians fell in between what they perceived their South Asian parent perceived as appropriate (i.e., not appropriate) and what they perceived their Canadian peers perceived as appropriate (i.e., quite appropriate). These data provide evidence that second generation South Asians see their views regarding sexuality as falling between two sets of cultural norms."

or the country of origin of their family, as in Bisin et al. (2008, 2016). Two reasons motivate this choice. First, behaviours are a better indicator of preferences than declared attitudes. Second, certain immigrants form identities with several layers. Attachment to the background culture or a non-Christian religion is not necessarily indicative of opposing the mainstream culture of the destination country.

The other main advantage of this outcome is that it is repeated, I observe it for every age between 15 and 20. This adds another dimension to the study. In addition to the outcome "do young women take more contraceptives if they live in an environment with fewer immigrants?", I can focus on the question "at what age do young women start taking contraceptives?". Neighbourhood characteristics can change the results on the finish line as well as the process which leads to it. This enriches considerably the picture that can be drawn from this paper. For instance, at age 20 women who grew up in neighbourhoods with a high concentration of co-ethnics could be similar to those who did not grow up in ethnic enclaves. However, they may have started taking contraceptives much later, at age 19 or 20, when their counterparts started at 15 or 16. Another scenario could be that the starting point is the same, there is no difference at age 15, but when women become adults, at 18-19, their paths start diverging. In both cases, this extra dimension would be informative on how neighbourhood characteristics affect cultural behaviours.

To be able to draw causal inference on neighbourhood characteristics, one needs to find an exogenous variation in residential choices. This is why, I focus on asylum seekers in the Netherlands. While their asylum application is being processed, asylum seekers are taken care of and hosted by a public organization, the Central Agency for the Reception of Asylum Seekers (COA). Allocation of asylum seekers into collective centres is done according to availability of places and not according to preferences. I use this institutional setting as a mechanism that quasi-randomly allocates asylum seekers across the country.

COA policy is to have their centres opened to the community they live in. This means children staying in centres go to local schools; centres are not closed, people can come and go in them. For these reasons, I treat allocation to a centre as equivalent to allocation to a neighbourhood. By looking at people living in centres, I focus on assignment to a point on a map, not to a polygon (as it would be the case if I looked at assignment to a city for instance). This means that I can zoom in into small neighbourhoods around the centre and still be safe that there is no residential sorting. This also allows me to compare the effects of closer/further away neighbourhood characteristics on the adoption of cultural behaviours.

As all papers, this one has strengths and weaknesses. The strengths are the exogenous variation (few settings allow to observe a random component of where people live) and the quality of the data. Thanks to the latter, I can both focus on very geographically disaggregated levels, calculate ethnic concentration in close neighbourhoods (of median size 2,647 people) while being able to track asylum seekers over a long period of time (the median age of arrival of the women whose behavior I observe between ages 15 to 20 is 6.8 years old). This gives credible variation in the close

environment in which one grew up.

There are two kinds of weaknesses, those that I can address and those that are inherent to the choice of usage of contraceptives as main outcome. Among the former, two main limitations come to mind. If this paper focuses on asylum seekers staying in temporary accommodation, how can one be sure that this allocation is meaningful? In the sense that people stay there for a long enough time? If any effect is observed, how can one dig into the mechanisms behind variation in ethnic concentration?

Sine I observe how long people live at each particular address, I can compare the effect of initial neighbourhood on those staying less and those staying more than the sample average. Since the mean duration is to stay 15 months in a centre, the subsample of people staying longer is one for whom the allocation was for sure meaningful. As for digging into mechanisms, the data is very rich and allows to look into many dimensions of treatment effect heterogeneity. There could be differences between the effect of ethnic concentration measured at the municipality and the (closer) neighbourhood level. There could be differences between the effect of ethnic concentration when all non-Western immigrants and when only those from the same origin are counted. There could be differences between arriving younger/older ... All of this can tell more about what is going on behind variations in ethnic concentration while being estimated with the same identification strategy this paper is built on.

Unfortunately, by focusing on contraceptives, I can only look at cultural behaviour of women. Nothing insures that the process for cultural convergence is similar between men and women and it is indeed one of the limitations of the paper. Another relevant result that I could observe for both mean and women, would be getting married to a native. Going this direction generates a trade-off. There is not enough people who arrived young in the Netherlands (say below 15) who are now in the age range 25 to 30 where one observes marriages. Increasing the sample size implies increasing the age at which asylum seekers arrive. However, focusing on a group who arrived below 15 is very different in terms of cultural convergence than looking at a group who arrived after age 15. Since the younger immigrants arrive, the closer they are to second generation immigrants (a relevant population for this topic); I choose to focus on contraceptives as an outcome and on asylum seekers aariving young as a population.

There are two main findings in this paper. First, with time, immigrants adopt the behaviours of natives. Those who arrived younger are more likely to have used contraceptives at least once by the age of 20. When pooling together all non-Western countries of origin, arriving one year older is associated with a decrease of 0.75 percentage points (for a mean outcome of 50% of usage). This means that someone who arrived five years older has 7.7% less probability of using contraceptives. This difference is particularly strong for immigrants arriving between 6 and 15 years old. Before that age, children are probably too young and an additional year is not associated with a statistically significant change in contraceptive usage.

Second, neighbourhood ethnic concentration does not have an effect on the adoption of natives

cultural behaviour. While the naive estimate, which does not account for residential sorting, shows a negative relation between contraceptive usage and size of the community from the same origin, the causal effect is a well estimated zero. The ITT estimate, where ethnic concentration is measured at the time of arrival, is very small in magnitude and statistically insignificant. The IV estimate, where concentration at age 15 is instrumented by concentration at assignment, is also insignificant. The results hold for both linear regressions (end point analysis) and duration models (adoption process analysis).

There is no evidence that this zero effect hides a significant result on a meaningful subpopulation. When ethnic concentration is interacted with a dummy for being below/above the sample (i) mean age of arrival, (ii) length of stay in a centre and (iii) centre size, none of the coefficients (neither those with an ITT or an IV interpretation, neither those estimated with linear regressions nor durations models) are statistically significant. These results ensure that the effect is zero for those who stayed longer than 15 months in the assigned neighbourhoods.

The rest of the paper is organized as follows: section 2 reviews the literature and details the contribution. Section 3 presents the data, while section 4 shows descriptive evidence on contraceptive usage by immigrants. Section 5 presents how the placement of asylum seekers can be used as a quasi-experiment and details the identification strategy. Section 6 presents the baseline results while sections 7 focuses on heterogeneity and potential mechanisms. Section 8 shows robustness checks and section 9 concludes.

2 Review of the literature and contribution

This paper lies at the intersection of two streams of literature. The closest literature, on cultural assimilation, matches topic-wise. The second closest, on ethnic enclaves, matches in terms of empirical strategy. In broad terms, I use the identification strategy of the ethnic enclave literature to study a research question from the cultural assimilation literature. To a lesser extent, this paper contributes to the literatures on (i) cultural change, (ii) neighbourhood effects and (iii) the usage of contraceptives.

Literature on cultural assimilation This paper is related to the vast literature on cultural transmission (Bisin and Verdier, 2000, 2001, 2010; Bisin et al., 2004; Patacchini and Zenou, 2016) and cultural assimilation (Olcina et al., 2017; Verdier and Zenou, 2017; Kuran and Sandholm, 2008; Bisin et al., 2008, 2011, 2016). The first set of papers focuses on the decision to transmit or not the cultural heritage. It distinguishes vertical transmission (from parents to children) from horizontal transmission (influence from the environment). A central question for the dynamics of these models, is the steady state characterized by a disappearance of the initial heterogeneity in cultural traits, is whether both types of transmission are substitutes or complements. The second set of papers looks at the decision to assimilate or stick to a distinct cultural background (Bisin

et al., 2011, 2016).

A simplified presentation would be that, in the first case, the decision makers are the parents (i.e. the first generation) while in the second the children take the decisions (i.e. the second generation)[3]. Because I look at the adoption of a behavior by immigrants who arrived very young in the Netherlands, the literature on assimilation is the most relevant for this paper. As a paper looking at both the descriptive evidence on convergence and neighbourhood effects, Bisin et al. (2008) find evidence of cultural assimilation being much slower for Muslim immigrants and no effect of neighbourhood segregation on the attitudes of immigrants.

My main contribution to this literature is to develop an identification strategy based on an exogenous variation in where people live[4]. My second contribution is to propose a consumption choice that can be labelled cultural: i.e. taking contraceptives when one is a teenager.

Literature on ethnic enclaves Recent papers studying the effects of living in more or less ethnically concentrated neighborhoods have relied on so-called "dispersal policies" implemented in the late 1980s, early 1990s mostly in Denmark (Damm and Dustmann, 2014; Damm, 2009) and Sweden (Aslund and Rooth, 2007; Aslund and Fredriksson, 2009; Aslund et al., 2003, 2011). The purpose of these policies was to spread out asylum seekers throughout the country once they were granted asylum status to avoid ethnic clustering. A central allocation office would decide where asylum seekers would live on the basis of only observational characteristics. That generated an exogenous variation that was used for identification. I rely on a similar natural experiment, following the work of (Beckers and Borghans, 2011) in the Netherlands.

I also focus on asylum seekers but at a slightly different stage of their settling in the destination country. The procedure for asylum seekers is similar in most countries and is made of two stages: first, after arrival, they are sent to temporary housing while their asylum application is being reviewed. In a second stage, if they are granted asylum status, they are free to go where they want. The studies in Denmark and in Sweden have relied on specific policies in the second stage. There is no such policy in the Netherlands but there is information available at the first stage that can be used for identification (for instance (Beaman, 2012) looks at professional network formation in asylum seekers centres and also uses variation in the first stage of settlement). Since asylum seekers do not chose in which centre they will live while waiting for being granted asylum status and, there is no evidence of sorting along cultural lines by the authorities, this setting can be used for identifying neighbourhoods effects.

As briefly mentioned in the introduction, the main difference in institutional setting between the Dutch and the Danish and Swedish cases is how it allows to look at assignment. The policies in

[3]Strictly speaking, in the literature on cultural assimilation, the decision maker culd be a first generation immigrant to be considered without family ties. Since, I focus on young immigrants, this is not the most relevant case.

[4]A noticeable exception is (Algan et al., 2013) who looks at whether immigrants give a foreign sounding name to their children, a question of cultural transmission more than adoption.

Denmark and in Sweden have randomly dispatched asylum seekers to municipalities. Everything that is averaged at the municipality level at the time of assignment can be considered orthogonal to the characteristics of asylum seekers. Nothing ensures that characteristics at the infra-municipal, for instance at neighbourhood level, are not endogenous[5]. In the Netherlands, assignment is made to a specific address, the one of the centre in which asylum seekers were temporarily living. The boundaries of the neighbourhood, whose characteristics can be considered orthogonal, are not fixed. What was assigned randomly is the centre of the circle, not the limit of the circle. This allows me to focus on neighbourhood characteristics at a very disaggregated level while still being confident that there is no sorting.

The literature has found positive effects of a larger community on labor markets outcomes of first generation, i.e. the parents, (Damm, 2009; Aslund et al., 2003; Beckers and Borghans, 2011), and educational achievements of the second generation, the children (Aslund et al., 2011). This paper highlights a potential backlash of theses positive results. Living in a more ethnically concentrated area may provide more opportunities for parents to find a job and a more stable environment, a more structured community for children to study. However, it also provides role models who are more distant from the mainstream culture. This paper focuses on the second part of the story.

A recent paper, Danzer et al. (2018), looks at the effect of ethnic concentration on the acquisition of language skills among second generation immigrants in Germany. It focuses on children of guest workers who arrived in the 1960s and 1970s as the authors argue that their geographical allocation was random. It finds negative effects of ethnic concentration. There are two main differences between this paper and mine. The first is the choice of outcome that cannot only be reduced to data availability. Language skills and contraceptives do not measure the same thing. While language skills are necessary to interact and integrate in the host country, it is a much less intimate measure and, one that is further away form what we can thought of as culture than contraceptives usage. The second is that I focus on a much smaller geographical level and thus am much more able to capture the effect of the environment where one lives.

Other Literatures My contribution to these literatures being much smaller, I go over them more quickly. I contribute to the literature on measuring cultural change and how it is influenced by peers. Recently, Bertrand and Kamenica (2018) studied how cultural distance between income/racial/gender groups evolved over the last four decades in the US. They find that cultural distance between groups is no larger today than it was in the 1970s. Fogli and Veldkamp (2011); Fernandez (2013) studied how Female Labour Force participation evolved over time in the US, building cultural change as a learning process where non working women may learn from already working ones from the same county. Although I differ from their approaches in many

[5]Although empirical evidence supporting non infra-municipal sorting has been shown in Sweden, Aslund et al. (2011)

ways (empirical methodology and population of interest) and that the results cannot be compared, my conclusion differs from theirs. I observe a decrease of distance over time and no effect of the environment.

I contribute to the literature on neighbourhood effects Kling et al. (2007); Chetty et al. (2016); Chetty and Hendren (2018a,b). These papers find that neighbourhood matters for educational and labour market outcomes only if children were young at the time they were living in these neighbourhoods. I do not find such heterogeneity in my results.

I also contribute to the development literature which studies contraceptives usage and fertility decisions. Munshi and Myaux (2006) shows that in rural Bangladesh, changing norms about contraceptives affect religious groups differently, although they have access to the same information. Kohler et al. (2001) show that the density of social networks have a different impact depending on the intensity of market activities. In both cases, contraceptive behaviours are influenced by peers behaviour.

3 Description of the data

I use two main sources of data: Dutch administrative registries collected and maintained by CBS (Centraal Bureau Statistik) and information on the location and operating dates of accommodations run by COA. CBS offers a very rich set of administrative datasets linkable through a unique individual identifier. This allows to put together information on various topics (medicine usage, location, family situation ...) and to link parents to children. A particular feature worth mentioning is that buildings in the Netherlands have a unique identifier which allows to identify exactly at which address everyone lives.

3.1 Usage of contraceptives

In the Dutch healthcare system, certain drugs are part of a basic package that every insurance company has to reimburse. CBS collects, on a yearly basis, which of these drugs is being dispensed to whom. Since it is compulsory to be insured, this data collects information on the entire population living in the Netherlands. Contraception, for women younger than 21 is part of this basic package[6].

More precisely, registries record categories of drugs according to the ATC4 classification. Data is collected from dispensed (and not just prescribed) medicine meaning that if a drug is prescribed but not collected, it does not appear[7]. This data is available for the years 2006-2017. For contracep-

[6] In 2011, a reform of the health system lowered the age from 25 to 21. To maximize the number of observations, I focus on the 21 age limit throughout the period 2006-2017.

[7] Usage of contraceptives appears in the data in the following specific cases: (i) women showing up at the pharmacy with a prescription that was already used in a previous year to buy contraceptives (or with an old tablet in the case of the pill), (ii) young woman showing up with a prescription (old or new) but wants to pay cash so that

tives, the most disaggregated entry I can use is the category G03A, "Hormonal contraceptives for systemic use". It includes the pill, patches, injections and implants, but not hormonal preparation used for the treatment of menstrual irregularities[8].

Drugs falling under the G30A category can also be used for the treatment of sever acne problems. In the descriptive part of the following empirical analysis, I compare, at a given age, girls from the same county of origin who migrated when they were younger/older. To make causal inference on neighbourhoods, I compare girls from the same country of origin who arrived in the Netherlands the same year but were allocated to more/less ethnically concentrated neighbourhoods. Likelihood of suffering from dermatological conditions is thus captured by age and country of origin fixed effects. Although the category G03A includes emergency contraceptives, the morning after pill is not part of the basic package and thus not recorded[9].

Contraceptives in the Netherlands can be prescribed by a GP and not necessarily by a gynaecologist. Parental consent is not necessary after a girl turns 16. It is possible, however, to ask doctors to wave this obligation for younger teenagers. According to the Personal Data Protection Act (Wbp), parents cannot access information on their children's treatments when they are older than 16. For a woman still living with her family, it is still possible that her parents find a tablet of pills for instance. It is a reminder that taking contraceptives at different ages should be interpreted differently. For instance, immigrant women from conservative families may be willing to start taking contraceptives as early as 16 y.o. but end up waiting to turn 19 or 20 y.o. by fear of being found out. Growing up in a less (more) ethnically concentrated neighbourhood can make teenagers more (less) willing to take the risk of opposing their families. This is why this paper uses heavily the fact that contraceptives is a repeated measure.

3.2 Data on COA accommodation and ethnic concentration

As of January 1^{st} 2018, the Netherlands is composed of 380 municipalities. Their sizes vary a lot, between 4,000 (Ameland) and 850,000 (Amsterdam) inhabitants. Since municipalities can be very large and are very heterogeneous, it is not be the best level to study neighbourhood effects. Instead, I focus on the "zip 4" level. A zipcode in the Netherlands is composed of 4 digits and 2 letters. The four digits divide the country in more than 4000 areas with a median population of 2647 inhabitants[10]. The entire zipcode ("zip 6") roughly corresponds to the street level.

no records of payment are kept (and potentially found by her parents).

[8]For more information, one can go to https://www.whocc.no/atc_ddd_index/?code=G03A&showdescription=yes

[9]G03A os itself subdivided into four categories, G03AA Progesterones and oestrogens, fixed combinations, G03AB Progesterones and oestrogens, sequential preparations, G03AC Progesterones and G03AD Emergency contraceptives. I do not have individual usage at this level but aggregate figures are available at https://www.gipdatabank.nl/databank#/g//B_01-basis/gebr/G03A for the period 2013-2017. The estimated number of G03AD users is around 3 out of 1000 of users of all G03A medicine. I take it as evidence that the data does not account for emergency contraceptives.

[10]In 2017, the country was made of 4,066 zip4 areas. The median population is 2647, while the mean is 3413 inhabitants.

There are two reasons why I mostly use the zip4 level to study the influence of the environment. First, it is not too small, not too big in the sense that it is large enough to approximate the sort of area in which someone lives and the people she encounters while not being so narrow (as "zip 6" would be) to miss a large share of the social interactions. Second, it is more uniform (in terms of population) across the country. Having information on the exact date when people are registered in and out of an address, I can calculate the ethnic composition of each municipality and zip 4 at the start of a calendar year.

I use the main population registry to tell whether residents are natives, 1st generation immigrants (born outside the Netherlands from non-Dutch parents), 2nd generation immigrants (born in the Netherlands from 1st generation immigrant parents) and to identify the country they relate to.

Another valuable information provided by CBS is (for first generation immigrants) the reason why they migrated to the Netherlands. CBS lists five main reasons for migration: family migration, asylum, work, studies and others. Information on asylum motives is well measured since the information comes from the Immigration Services (IND)[11]. It can be used to distinguish between economic migrants and asylum seekers among immigrants who arrived from the same country the same year[12].

I combine this administrative data with information on all accommodations listed by COA between 1995 and 2015. As will be explained at length below, asylum seekers are under the responsibility of COA during the time when their asylum application is processed. During this period, they can live in a collective centre together with other asylum seekers or with relatives (or friends) already living in the country. In any case, the address where they are staying is know by COA. I have access to all these addresses. This gives a total of 17,000 different addresses used by COA for the period 1995-2015 out of which, 15,500 could be coded for by CBS and merged with municipal registries.

3.3 Control variables and neighborhood characteristics

I observe the age, gender, marital status of parents and the number of children they had at the time of migration. The information on marriage is available for residents who got married before arriving to the Netherlands as they need to register their union.

[11] However, one has to be careful with migrants listed under family migration. Many asylum seekers arriving as children are listed as family migration if they did not arrive the same day as their first relative who entered the country (although the latter category is listed as asylum seekers).

[12] Although, one rightfully expects certain combinations of country and year to produce a large majority of asylum seekers (as compared to regular immigrants), it is not straightforward to classify all immigrants from a country/year combination to be asylum seekers or regular immigrants. For instance, when focusing (i) on the ... countries that produce most of the asylum seekers followed in the causal part of the paper and (ii) on the years in which those countries supplied more than 100 immigrants, 16% of the country/year combinations had a ratio of asylum seekers to asylum + work + studies migration motives between 25 and 75%, meaning not supplying either almost exclusively or almost no asylmum seekers.

As an additional individual control, I have information on the highest level of schooling recorded for people living in the Netherlands. This data usually contains many missing values for first generation immigrants. Adult immigrants did not complete their education in the destination country and thus often do not appear in registries. A distinguishing feature of the data source used by CBS is that it combines different sources of information on education levels to select the highest one. For instance, information is gathered from high schools, universities ... and the unemployment agency. The last one increases dramatically the number of asylum seekers for whom I have information on education level. This is very helpful when I perform balancing tests. It does not report, however, when (and where) the highest education was acquired.

The objective of this paper is to measure the effect of the quantity of people from the same ethnic background. It is important to check that the effect is driven by ethnic concentration and not "quality" of the neighborhood (high unemployment, high crime rate ...). This can be a co-founder as it could be harder, for instance, in low "quality" neighbourhoods to have access to a doctor or a pharmacy. I use information on the "quality" of zip4 areas provided by the Netherlands Institute for Social Research, the *status score*. Every four years, this government agency produces a ranking of all the 4 digit zipcode areas based on the average income in a neighborhood, the percentage of people with a low income, the percentage of low-educated people and the percentage of people who do not work. By means of a factor analysis (principal components analysis), these characteristics are summarized in one composite characteristic: the *status score*. I use the normalized measure provided by the Institute for Social Research for the years were they are available and fit lines between these years to predict the missing observations. This measure of neighbourhood quality has the benefit of being time-varying, which is relevant when one focuses on long periods of time. The normalized measure, extended for all years, show an intergroup correlation of 0.78.

3.4 Sample restrictions

I focus on complete spells (women with observations for all ages from 15 to 20). This is because I am interested in the "final" outcome, has adopted or not the behaviour of interest, as much as in the adoption process itself. Consequently, I look at women who turned 15 in 2006 up until those who turned 20 in 2017, so who were born earliest in 1991 and latest in 1997. Since I want to observe women at age 15, this is the oldest they can arrive in the Netherlands. THis means that they migrated at the latest in 2011.

As a consequence, I cannot use measures of behaviours (other women taking contraceptives) as the main explanatory variable. This information is only available from 2006 and relying on it would greatly decrease my sample size. Instead, I use ethnic concentration. Restricting to complete spells also makes sure that women have spent some time (at least five years between 15 and 20 y.o.) in the Netherlands. This way, I make sure that I do not capture a mechanical effect of access

to healthcare[13] but one of cultural assimilation. I call this sample the *complete spell* sample.

Because the data on contraceptives is available from 2006 onwards, certain women arrived before age 15 in the Netherlands and turned 15 before 2006. They are not in the *complete spell* sample. I create a sample of complete and incomplete spells with all women who arrived before age 15 in the Netherlands and whose contraceptive usage is observed until age 20. This will be used in the robustness checks.

4 Descriptive Evidence on Convergence in Behaviours

I first detail how usage of contraceptives differs between natives and immigrants and then estimate if there is evidence of convergence: with time do immigrants behave more like natives? Do immigrants who have been in the Netherlands for longer are more likely to take contraceptives? To do so, I compare (i) first and second generation immigrants of the same country and then (ii) compare first generation immigrants of the same country who migrated in the Netherlands at different ages. Depending on how old they were when they arrived, certain immigrants would have been in the country for longer when they reach their teenage years.

4.1 Differences between natives and immigrants

Table 1 reports the probability for natives, first and second generation non-western immigrants to have purchased contraceptives at least once at all ages between 15 and 20 years old. The first three columns show the number of girls in each demographic group, the last three report usage.

Table 1 here

Three main elements emerge from this table: first, there is a striking difference between the consumption behaviour of natives and (first and second generation) immigrants. While at age 20, almost all young native women (87%) in the Netherlands have taken contraceptives at least once, only half of first generation immigrants have. Second, disparities are sharp at every age. It is indeed largest at younger ages, for instance at age 15 (when parents approval is necessary), the rate of consumption is more than twice larger for natives than for second generation (11 vs 24%) and more than three times larger for natives than first generation immigrants (7 vs 24%). The differences remain throughout the spell, with natives having a rate of usage at least twice larger than first generation immigrants until age 19. Finally there are (limited) differences between the behaviours of first and second generation immigrants: second generation immigrants behave more like natives than first generations.

[13]To make sure of that, I check what percentage of women followed in the descriptive and causal parts of the paper can be found in the drug database. This means, what proportion collected a drug from the basic package in the period 2006-2017. Respectively for the populations followed in the descriptive and causal parts of the paper, 98% and 95% appear in the medicine data, meaning they do have access to healthcare services.

To investigate if these differences are due to a change in the composition of origin countries between first and second generation teenagers between 2006 and 2017, I reproduce the same table for the five non western countries that supply most of the immigrants shown in table 1, namely Turkey, Morocco, Surinam, Indonesia and Yugoslavia..

Table 2 here

The picture is very heterogeneous. There are clearly two groups of countries: Turkey and Morocco on the one hand, Surinam, Indonesia and Yugoslavia on the other. While rates of consumption for the latter countries are within the same order than those of native Dutch, (73, 82 and 70% for second generations compared to 87%), Morocco and Turkey are largely behind with consumption rates at age 20 that are below natives consumption rates at age 16. Another striking element is that second generation immigrants from Turkey and Morocco do not seem to behave very differently from first generation immigrants. There is a maximum 5 percentage points difference between these categories, i.e. Turkish immigrants at age 19. Discrepancies are larger for the other group of countries, as high as 20 percentage points for immigrants from Yugoslavia at age 19.

4.2 Differences in behaviour over time

Comparing first and second generation immigrants from the same country in the same years does not provide fine evidence on the evolution of behaviours. Any difference can follow from a change in the composition of immigrants from the same country through time rather than from the adoption of natives behaviour. To further analyse whether there is convergence, one wishes to see if immigrants who have been longer in the Netherlands tend to behave more like natives. The natural way to proceed is to regress behaviour on the length of stay in the country. I do so, in a slightly different way. Since I observe the outcome at all different ages between 15 and 20, I include age at migration as a regressor in separate regressions for every age between 15 and 20. More precisely, I run the following regression:

$$y^a_{i,c,t} = \alpha + \phi(\lambda_c, \theta_t) + \beta\, age_t + \epsilon_{i,c,t} \tag{1.1}$$

Where $y^a_{i,c,t}$ is a dummy for having taken contraceptives at least once by the age of a (for $a \in \{15, 16, 17, 18, 19, 20\}$) for individual i from country c who arrived in year t in the Netherlands. λ_c are country of origin fixed effects and θ_t are year of arrival fixed effects. The variable of interest is age_t the age (expressed in years) at which individual i arrived in the country. $\phi(\lambda_c, \theta_t)$ indicate that I will control for year of arrival and country of origin through various combinations of fixed effects. I try three specifications $\phi(\lambda_c, \theta_t) = \lambda_c$, $\phi(\lambda_c, \theta_t) = \lambda_c + \theta_t$ and $\phi(\lambda_c, \theta_t) = \pi_{c,t}$ where $\pi_{c,t}$ refer to fixed effects interacted for year of arrival and country of origin.

Equation 2.1 does not include the year in which individual i turns age a, since it is a linear combination of θ_t, a and age_t. It would limit the variation used to identify β to differences in months of arrival (since age_t is calculated using birthday and day of entry in the Netherlands and is not an integer). This corresponds to the behavioural assumption that, for instance, taking the pill at age 17 is similar in different years between 2006 and 2017[14].

In addition to equation 2.1, I pull together the observations for all the ages and estimate a Weibull MLE using the fact that starting to take contraceptives can be defined as a failure in a duration analysis framework. The hazard rate $\xi(t)$ has the following form, where t is the age between 15 and 20 at which a woman starts taking contraceptives and ρ is a shape parameter determining whether the failing rates increase or decrease with time:

$$\frac{f(t)}{1-F(t)} = \xi(t) = \rho \, \exp(\alpha + \phi(\lambda_c, \theta_t) + \beta \, age_t + \epsilon_{i,c,t}) \, t^{\rho-1} \tag{1.2}$$

The coefficient on age should be read as the percentage increase in the hazard rate from arriving one year older in the Netherlands. Results of the estimation of equations 2.1 and 1.2 are reported in table 3.

Table 3 here

There is evidence of convergence in behavior. Arriving older decreases the probability of behaving like a native. This effect can be seen with both the linear specification and the duration model. The effect is robust to various ways of controlling for country of origin and year of arrival fixed effects. The effect of arriving younger is quite sizeable and is more easily interpretable with linear regressions. The effect of arriving older is sharper at earlier ages (compare to unconditional mean) but remains throughout. At age 20, two girls arriving the same year from the same country, one being five years older than the other has 3.75 percentage points less probability of having ever taken contraceptives. It gives support to the idea that immigrants (at least those who arrive young, before 15 years old) adapt their behaviour to the mainstream norm the longer they have been in the destination country. Having negative results at all ages with linear regressions and when using the Weibull specification shows that the final outcome at age 20 but also the path leading to change with age at arrival. These numbers can also be used as a benchmark for the causal part of the paper since the effect of ethnic concentration can be compared to the effect of arriving one year older.

There are two main caveats to this descriptive analysis. First, arriving younger and thus being exposed to the Dutch mainstream culture for a longer period also means arriving at a younger

[14]In table A1 in appendix, I test whether the environment is "stationary" in the sense that $\mathbb{E}[y^a|\lambda_c]$ is the same across the years in which women turn a in my sample. To do so, I regress y^a on λ_c and years when women turn age a and test the joint null that all the years have an equal mean. $\mathbb{H}_0$ fails to be rejected at conventional levels for ages 19 and 20. Even for ages where $\mathbb{H}_0$ is rejected, differences between years are small in magnitude when reported to the level at the reference year (for ages older than 15).

age when one's attitudes are potentially more malleable. To make sure that the difference is attributable to children, at a similar stage of their development but who have been immersed to the destination country for longer, I re-estimate equations 2.1 and 1.2 but interact age_t with three dummies taking value one if a child arrived before age 5, between 6 and 10 and between 11 and 15. Results are reported in table 4[15]. The effect is still very strong at age 20 for children who arrived after 5 y.o. and left almost unchanged in the duration model.

Table 4 here

The second caveat could be that the estimation does not allow the effect of age at arrival to differ by country of origin. To circumvent this problem, I re-estimate equations 2.1 and 1.2 on a subsample of the five largest suppliers of first generation immigrants, namely Turkey, Morocco, Surinam, Afghanistan and Iraq and interact age_t with origin country dummies. The results are shown in table 5 for linear regressions.

Table 5 here

An interesting element emerges from table 5. While there is evidence supporting convergence for immigrants from Afghanistan, Iraq and Surinam, there is none for those from Turkey and Morocco. Results for immigrants from Iraq and Afghanistan are larger than the corresponding unconditional coefficients from table 3. It shows that the heterogeneity is not limited to Muslim vs non-Muslim origin but that there are different rates of convergence within immigrants from mostly Muslim countries. This can be seen as an additional motivation to focus on the role played by ethnic concentration. Among majority Muslim countries, those who show the lowest rate of convergence are also those with larger existing communities in the Netherlands (Morocco and Turkey).

This section has established two facts that make usage of contraceptive a good measure of cultural behaviour: (i) they are differences in consumption between immigrants and natives (there is something to be explained), and (ii) there is evolution in immigrants behaviour through time, this behaviour is not fully inelastic and it is relevant to see if the environment influences it.

5 Institutional Settings and Identification Strategy

Once the descriptive evidence on the convergence in cultural behaviour has been established, it is interesting to dig into what can increase or slow down its pace. One contender is the ethnic concentration in which young immigrants live. It is difficult to gather evidence on the subject since one must rely on some exogenous variation in residential choices to assess the importance of

[15]Results for duration models when $\phi(\lambda_c, \theta_t) = \pi_{c,t}$ are very computationally demanding and available upon request.

neighbourhoods. Otherwise some unobserved individual characteristic, for instance family conservatism could both drive the decision to live in certain neighbourhoods (with a larger/smaller ethnic community) and influence the decision of the children to use contraceptives. Below I describe the ideal experiment one would run to estimate this effect and show how the setting of asylum seekers in the Netherlands relate to this experience.

The ideal large-scale real-life experiment would be to take young girls whose cultural background differs from that of natives, allocate them randomly in different environments, let them grow up in the assigned places until they become teenagers/young adults and then observe their behaviours.

The natural experiment used in this paper fulfil these requirements. I observe young girls who arrived in the Netherlands as asylum seekers. They arrived (on average at age ...) from countries whose cultural background is arguably different from the Netherlands (mostly from Afghanistan, Iraq, Iran, Yugoslavia, Somalia ...). While their application is being processed, they are taken care of by a public organisation in charge of welcoming asylum seekers, the Central Agency for the Reception of Asylum Seekers (COA). They are in charge of accommodating asylum seekers in one of the 200 or so centres they have in the country. Since asylum seekers do not choose where they go and COA does not cluster them by how conservative they are, this first assignment provides exactly the exogenous variation required to make causal inference.

Note that in this setting, asylum seekers are allocated to a point on the map and not to a larger polygon (for instance a municipality). This allows to treat the immediate surroundings as exogenous contrary to settings where asylum seekers are sent randomly to a city. In the latter case, neighbourhoods characteristics have to be averaged at a pre-determined (and much higher) level. I later observe the behavior of women between ages 15 and 20.

This setting departs from the ideal experiment in that asylum seekers do not stay in government provided housing indefinitely. At some point (on average after 15 months), they are out of COA responsibility and move to traditional housing. To make sure that the setting is still valid, I look at two things in particular: (i) how long asylum seekers stay in the neighbourhood (and municipality) in which they were first assigned? This provides evidence that assignment meaningfully influences future residential choices, (ii) how heterogeneous are baseline estimations to length of stay? Is the effect different for asylum seekers who stayed in COA centres for a long term (more than the mean duration of 15 months)?

5.1 Institutional Setting

Asylum placement procedure

Asylum seekers arriving to the Netherlands by plane or at another point on the border can claim asylum once in the country. The first stage of the application process is to go to the Immigration and Naturalisation Service (IND) facility in the village of Ter Apel in the North of the Netherlands.

After been giving six days to recover from their trip, asylum seekers are first interviewed by the Aliens Police to check their identity and their motives for seeking asylum. They can also see a doctor.

At the end of these few days, asylum seekers can either (i) be denied further stay in the country, (ii) be granted refugee status or (iii) it can be decided that their application needs more time to be processed. In the first case, they should leave the county. The second case, almost never happens. In the last case, asylum seekers are placed under the responsibility of the Central Agency for the Reception of Asylum Seekers (COA) which is responsible for accommodating them. I exploit this first allocation as an exogenous variation in where asylum seekers live.

At any time, COA manages centres throughout the country. When a family needs to be hosted, COA looks for a centre that can accommodate it. The allocation is done centrally by COA and the municipality where the asylum seekers will live has no say in this allocation. Since 1996, COA is responsible for the housing of all asylum seekers in the Netherlands. During the period when their asylum application is being reviewed, asylum seekers have to live at the address where they are being registered at COA. The main criteria for COA to allocate asylum seekers is availability of a place for a given family composition.

There are two obvious threats to the exogeneity of this allocation that I discuss in details below. The first one is that the asylum seekers could choose to live in a specific centre (and in the environment that goes with it) or that COA could send specific families to specific locations. Although possible on paper, this scenario seems unlikely in practise. From discussions with COA personnel, the main difficulty when allocating families is to find a centre ready to host a family of a certain size when it arrives in the country. The difficulty comes from the limited supply of housing (that pre-existed the 2014 spike in asylum applications). The only case in which asylum seekers (or COA) could express a preference, would be if several centres had the capacity to host the same family at the same time. It is unfortunately not possible to isolate these cases in the data but it was not pointed as a frequent scenario in discussions with COA. To alleviate any related concern, I perform numerous balancing tests to back up the information about the institutional setting.

The second threat is the existence of family ties in the Netherlands. Asylum seekers with relatives or friends already living in the Netherlands could use the fact that they have pre-existing ties to influence the allocation process. This is a real concern since asylum seekers have the possibility to live with their relatives. However, this case is very well documented in the data. If relatives want to live with family members, they have to register their address at COA. They are thus easily identifiable. Since this allocation cannot be considered as exogenous, I only consider asylum seekers living in collective housing (and not at individuals).

To observe sorting, families who do not have relatives (or friends) in the Netherlands (at least not ready to welcome them) would need to arrive at a time where several housing opportunities, suitable for their family size, are available when it is usually pointed out that supply is very limited. They would also need to have information (at a detailed geographical level) on neighbourhood

ethnic composition (the main explanatory variable) to decide to sort along that dimension. It does not seem ex-ante to be a very frequent and likely scenario. A doubt however can persist on the exogeneity of the allocation. This is not because there appears to be a lot of latitude for asylum seekers to decide where they will live but because there are no definite guidelines on how the allocation is made. This is the reason why these doubts have to be settled by more formal statistical tests.

Asylum seekers can ask to be relocated to another center if for instance they have first grade family members in another center of if they have an employment opportunity somewhere else. It can also be that some centers close down and asylum seekers are sent to another center somewhere else. The latter case is arguably random, the former not. To make sure that I capture an exogenous variation, I focus on first placement.

After being granted refugee status, certain asylum seekers stay in COA centres while looking for an accommodation outside the centre. COA can help them in that process and tries to find a longer term accommodation in the same municipality. Refugees at this stage are however free to go wherever they want. As for unaccompanied minors, those between 15 and 18 live in a specific type of centre. Those who are below 15 can live with a foster family. If they do not, they also stay in a specific type of centre.

How to identify the experimental population?

To identify where asylum seekers live and whether this location can be considered as an exogenous assignment, I combine information on the exact addresses and operating periods of all COA registered accommodation (meaning centres and addresses of relatives) for the period 1996 to 2016 with municipality registries. COA operates various types of centres (for adults, for minors ...) and addresses of individuals can be registered under different categories. Table 6 lists the type of accommodation for which I have an address that could be merged with CBS data.

Table 6 here

The same address can be listed under different types of centres, for instance a regular collective centre, an *Asielzoekerscentrum*, can also be listed as one that welcomes unaccompanied minors, a *Kleinschalige Centrale Opvangeenhdn* if a wing of the building is used specifically for minors. The second column shows how unique addresses are distributed among the different types of accommodation. Columns 3 to 5 show the same distribution respectively for addresses that appear twice, three or four times.

The main distinction between the centres are whether they are individual or collective housing. As pointed out earlier, individual housing are mostly asylum seekers being hosted by relatives already living in the Netherlands. Asylum seekers whose first placement is at one of these addresses are not part of the experimental population. This excludes *Zelf Zorg Arrangement*, *Aanvullende Opvang* and *Gemeentewoning* from consideration.

Another way of distinguishing types of accommodation that correspond to individual and collective housing is to look at the number of addresses listed by category. If a type appears very often (say more than 250 times), it is more likely to capture many different individual situations of people living at different addresses than it is to capture a collective housing. It is reassuring that this (data driven, more neutral) approach gives the same conclusion than following institutional information (but for the category *Kleinschalige Centrale Opvangeenhdn*).

To identify asylum seekers, I look at non-Western[16] immigrants registered in the building listed by COA. I do not consider potential asylum seekers living in the same 6 digit zipcode level (i.e. the street) than an operating centre. This is a very conservative approach since several COA centres are large and potentially spam over several addresses. I also use the exact day of registration and opening and closing of the centers which ensures that a particular address is a COA center at the time asylum seekers moved in.

To identify my experimental population, i.e. the asylum seekers that have been spread out randomly over the country, I use the following criteria: (i) the first address at which an asylum seeker was registered has to be one of a collective housing centre (in use at the time)[17], (ii) if an address is listed both as collective and an individual type of accommodation, I use the conservative approach of not counting this observation in my experimental population[18].

5.2 Additional information

In typical asylum centres inhabitants take care of themselves as much as possible. They are responsible for taking care of their living environment. They are free to go outside of the centre and are stimulated to do so: they go shopping (they cook themselves), they can participate in local sports clubs, their children go to local schools, they can do volunteering work outside of the centre. Every week they receive money for food and clothing depending on the composition of their family and their income. In addition, they can work at the centre for a small payment. After 6 months, they can also look for a regular job. So by no means are COA centres closed and indeed living there means interacting with the local community.

[16]Non-Western is being defined as immigrants not from Northern Europe, Western Europe, North America and Oceania according to the classification of table 12.

[17]A certain number of observations appear as being listed first in a non COA registered accomodation (neither single, nor collective housing) for a small period of time and then in a collective housing. If an individual spent less than one month in an unregistered location and then joined a collective housing, it is included in the analysis.

[18]I perform a last check to see if asylum seekers can be identified from municipal registries (where individuals are linked to an address). Since 1996, COA is responsible for the placement of asylum seekers, so all asylum seekers should be listed at least once as living in COA registered accommodation (meaning an active address at the time). I use the information provided by IND on migration motives to see whether all people listed as asylum seekers can be found that way. 75% of all people can be traced back with this approach (if one focuses on building identifiers) and 84% can if one extends to looking at zip6 with an active address. This is a very high number considering that only 85% of the addresses could be identified and anonymised by CBS.

Is assignment a good anchor for later residential choices?

The setting used in this paper differs from previous work that has relied on the exogenous assignment of asylum seekers in that I look at assignment into temporary accommodation. It is important to check that although temporary, this allocation is meaningful: asylum seekers should spend time in the zip4 they were assigned to. Columns 2 and 4 of table 7 report the mean and main quantiles of the distribution of length of stay in centres. Column 2 reports these informations for all women with complete spells who are part of the experimental population. Column 4 reports the same information but for complete and incomplete spells. The support of the distribution is very large, some people spending just a few weeks while others almost three years. I choose not to drop observations below a certain length as to not manipulate unnecessarily the data. Two facts are reassuring: first, the central measures of the distribution, i.e. the mean and the median show that the assignment is meaningful, 9 months for the median, 15 months for the mean. Secondly a substantial number of individuals spent a long time in the centres (the 75 percentile is more than 21 months) so that there is a (large) subsample for which assignment converted into long treatment.

Table 7 here

In table 8, I calculate the number of asylum seekers still residing in the Netherlands 1 to 8 years after having arrived in the country. I also compute how many live in the same 4 digits zipcode than the one they were assigned. One constraint is that the zip4 level is very small and expecting people to live so close to the centre is probably too restrictive. This is why I also show how many live in the same municipality. Since that level is very large (especially for big cities), I try to find a middle ground by geocoding all the adjacent 4 digits zipcode areas. This creates a series of larger (and overlapping) entities with median population around 20 000 inhabitants. After 3 years[19], 21.4% of the assigned girls are living in the same municipality. Only 13% live in the same zip4 (most likely still living in the centres where they were assigned). 40% of the people living in the same municipality after three years are thus no longer in centres. This could be driven by people living in large municipalities and moving from two (distant) points within the same municipality. This does not seem to be entirely the case, since more than half of those who live in the same municipality but not in the same zip 4 live in an adjacent zip4. This shows that the assignment is indeed significant in two ways: (ii) it influences future residential choices and (ii) although asylum seekers did not stay forever in these centres, many stayed for a long time .

Table 8 here

[19]I calculate three years from the starting date of registration in the centre, i.e. see whether asylum seekers still live in the same zip4/adjacent zip4/municipality 3*365 days after ben assigned to a centre.

Is there variation in the ethnic composition of the neighborhoods?

I distinguish two moments in the analysis, when asylum seekers are assigned to a location and when they turn 15 years old. I look at two definitions of ethnic community, one where I group countries of origin into various regions, as classified in table 12 and one where I collapse all non Western immigrants[20]. In table 9, I report the main quantiles and the mean of the distributions of community sizes for both definitions of ethnic groups, at assignment and at age 15, at the zip4 and at the municipality level.

Table 9 here

There is a lot of variation to be exploited even at the zip4 level. The 75th percentile is more than five times larger than the 25th. It also gives an idea of what it means to increase the size of the ethic community by 50 or 100%, a critical element when interpreting the results. To give a more detailed view of the heterogeneity (by origin country) in the magnitude of the variation in community size, I reproduce table 9 for the five largest countries of origin of the experimental population (see tables A3, A4, A5, A6 and A7 while table A2 shows the distribution of origin countries).

In figure 1, I plot the log of ethnic concentration at age 15 against the log of ethnic concentration at the time of assignment. In this figure ethnic concentration is defined as immigrants from the same region and neighbourhood refers to zip4. In figures A1, A2 and A3, I reproduce the same graph for the other combinations of neighbourhood and ethnic community. Each figure is split into four quadrants, in the upper left one, I scatter the points, in the upper right one I approximate the scatter of points with a locally weighted regression, in the lower left one I use a quadratic approximation and in the lower right one a kernel-weighted local polynomial. Although some of these approximations are highly non parametric, they show a monotonic relation between concentration at assignment and concentration at age 15.

Figure 1 here

5.3 Balancing tests

As hinted in figure 1, I follow the literature (Bertrand et al. (2000); Aslund et al. (2003); Damm (2009); Aslund et al. (2011)) and measure ethnic concentration by taking the log of the size of the community. Using the size of the ethnic community rather than its share over total population has one particular advantage: although zip4 are more similar in sizes than municipalities, they are still relatively heterogeneous. The same measure of share, say 2%, could be 20 over 1000 or 40 over 2000 and it is not the same to live close to 20 or 40 co-ethnics. The drawback of this

[20]Meaning not from countries in North America, Northern Europe, Western Europe and Oceania. I pool together 1st and 2nd generation immigrants not living under COA care together with people who are living in COA facilities.

is that countries of origin have to be pooled into regions (see table 12) to avoid having zeros for individuals with small communities in the Netherlands.

The balancing tests work as follows: I separate variables that should enter the assignment of asylum seekers, namely demographic characteristics (Z_i) from those that should be unrelated (X_i). To test the exogeneity of assignment, I see whether ethic concentration at the time of arrival is related to Z_i and X_i.

To do so, I regress $\ln e_{i,t,h}$ the log of the size of ethnic community of individual i who arrived in year t in neighbourhood h on the characteristics of the household head (identified at the time of arrival of the young girl[21]). It includes the following variables as X_i and Z_i in the estimation of equation 1.3, gender (if the head is male), age at migration, the number of children at arrival and dummy variables for education level[22], keeping in mind that this information is available for 85% of household heads in the sample. I estimate the following equation:

$$\ln e_{i,t,h} = \alpha + \beta \underbrace{X_i}_{Unrelated} + \gamma \underbrace{Z_i}_{Related} + \lambda_c + \theta_t + \epsilon_{i,t,h} \tag{1.3}$$

where all specifications control for country of origin and year of arrival (since they are the default controls in the main regressions). Results for concentration at the zip 4 level are reported in table 10, those for concentration at the municipality level are reported in table 11. I report the results in columns labelled (1) and (3) where in (3), I add municipality of assignment fixed effects. I also report the F-test of equality of all the education dummies together with the p-value associated with this test. It is clear from looking at columns (1) and (3) in both tables 10 and 11 and for both definitions of ethnic community (from same origin and all non-Western immigrants) that neighbourhood characteristics are only correlated with demographic conditions of the family to be hosted. There is absolutely no evidence of sorting, in the sense that more educated asylum seekers are no less likely to live in an ethnically concentrated area than less educated asylum seekers. P-values for sorting with immigrants from the same origin at the zip4 level, the main worry are extremely high (above 0.95).

Tables 10 and 11 here

To give more credibility to these tests, I show that they have power against the alternative of sorting. To do so, I regress ethnic concentration in the neighbourhood where young women live when they are 15 on characteristics of the household head at that time (meaning the household head age when their daughter turns 15, the number of children that year ...). Education levels

[21]The head is identified as the father of the girl if he arrived at the same time, otherwise it is the mother. Unaccompanied children are not associated with an head and thus are not included in the estimation of equation 1.3.

[22]I create five dummy variables, missing information, less than primary school (basic education), primary school, middle school, high school and higher education. The distribution in the main sample is the following: : 31%, 22%, 27%, 10%, 10%.

remain the same since the data is silent on when education was followed. To make sure that individuals do sort into neighbourhoods, I focus on those who do not live at age 15 in the same zip4 as they were assigned. I thus run the same regressions at age 15 that I did at assignment for the balancing. I do it though on a population and at a time where there has been room to select where to (and close to whom) live. In this context,the regressions at age 15 should reject no sorting.

Results are reported in columns labelled (2) and (4) in tables 10 and 11 with (4) including municipality (at age 15) fixed effects. The picture is very different from the balancing tests. There is clear evidence of sorting. In particular, highly educated people live in zip4 with smaller ethnic communities (26% smaller when estimating without municipality fixed effects, 10% when including them). These tests suggest that sorting is stronger along people from the same origin and for close neighbourhoods, the expected pattern. Being able to unequivocally reject the null of no sorting in (2) and (4) give credit to the results in (1) and (3) and establish the exogeneity of first assignment.

6 Empirical Analysis

6.1 Specifications

All specifications run in this paper are characterized by four elements. Each of the four dimensions boils down to a choice regarding one of the variable in the estimated equation. I present the various choices, then I show the estimated equation and finally discuss specific points on the identification and estimation.

First, the outcome variable. The choice is between using one observation per individual, namely a dummy variable that takes value 1 if a woman has ever taken contraceptives by the age of 20, or using multiple observations per individual, a dummy has ever taken contraceptives by each age between 15 and 20 years old. The first case selects some point in time (the latest possible with the data available) and ask the question did we observe influence of the environment at this end point? It can be estimated by cross section linear models. The second case answers the question: does the environment change the path leading (or not) to convergence? In a fashion similar to the descriptive part, all specifications are run twice, with linear and non linear regressions.

Second, the main explanatory variable, the log of co-ethnics. Ideally, I would like a measure of concentration at age 15, when usage of contraceptives start being observed, and I would like it to be exogenous. Failing of that, I pursue three types of analysis: (i) the naive one in which I use concentration at age 15 as the variable of interest, (ii) an ITT strategy in which I use the neighbourhood ethnic concentration at the time of arrival and (iii) an IV strategy where I instrument concentration at age 15 with concentration at the time of arrival[23].

[23]The terms control functions and IV with duration models are used interchangeably as control functions can be seen as non linear instrumental variables Wooldridge (2015).

Third, the choice of controls. For all the models that I estimate, I use five different sets of controls, I number them (1) to (5) and use this numbering in the tables to specify what controls are used in which columns. (1) refers to country of origin and year of arrival fixed effects, (2) adds age at migration of the girl to (1), (3) adds individual characteristics of the household head, namely age at arrival, dummies for education level, gender and nb of children at assignment for ITT and IV and at age 15 for naive, (4) adds the index value on neighbourhood "quality" to (3) and, (5) adds municipality, at assignment for ITT and IV and at age 15 for naive, to (3).

Fourth, should ethnic concentration enter linearly? or should one expect its role to be quadratic? Currarini et al. (2009) showed that the patterns of friendship formation differ between small and large groups on the one hand and medium sized groups on the other. This points to a quadratic effect of ethnic concentration. Loken et al. (2012) has shown the potential pitfalls of estimating linear models when the underlying relation is not. This motivates adding a quadratic model to the linear one. This means that there are two models to estimate: the linear one,

$$y_{i,t} = \alpha + \beta \ln e_{j,g,t} + \pi X_i + \epsilon_{i,t} \tag{1.4}$$

and the quadratic one,

$$y_{i,t} = \alpha + \beta \ln e_{j,g,t} + \gamma (\ln e_{j,g,t})^2 + \pi X_i + \epsilon_{i,t} \tag{1.5}$$

where $y_{i,t}$ is a dummy, X_i are the set of controls detailed, $e_{j,g,t}$ is the relevant community size, j referring to the geographical area (zip4 or municipality), g refers to groups , i.e. from the same region or all non-Westerns pooled and t refer to age 15 for the naive estimate and to assignment for the ITT and IV strategies[24].

The identifying assumption of the ITT strategy is that conditional on year of arrival and country of origin, ethnic concentration of the neighbourhood is exogenous. This means that an individual who arrived from a certain country in a certain year could be sent to different types of neighbourhoods, i.e. there was no reason why he or she would be sent to one in particular. Do note that this identifying assumption allows for clustering of individuals from the same country as long as specific individuals from the same country are not sent to specific places.

In addition to the assumptions just listed, an instrumental variable needs to satisfy relevance (ethnic concentration at age 15 is significantly related to concentration at arrival) and the exclusion

[24] I only show linear regression models for simplicity, the corresponding duration models are .

$$\frac{f(t)}{1-F(t)} = \xi(t) = \rho \, \exp(\alpha + \beta \ln e_{j,g,t} + \pi X_i + \epsilon_{i,t}) \, t^{\rho-1}$$

for the linear model and

$$\frac{f(t)}{1-F(t)} = \xi(t) = \rho \, \exp(\alpha + \beta \ln e_{j,g,t} + \gamma (\ln e_{j,g,t})^2 + \pi X_i + \epsilon_{i,t}) \, t^{\rho-1}$$

for the quadratic one. For simplicity, I refer to equation 1.4 and 1.5 for duration analysis too.

restriction (ethnic concentration affects the decision of taking contraceptives only through its influence on ethnic composition at age 15) to be valid. To be given a LATE interpretation, it must also satisfy monotonicity (a high/low ethnic concentration at arrival means a higher/lower concentration at age 15). Under these assumptions, the effect identified is the change in behaviour for those who live in a higher/lower (depending on the sense of the monotonicity) ethnic neighbourhood because they were assigned to a high/low ethnic neighbourhood when they arrived.

Although figures 1 and A1 show evidence supporting monotonicity, there can be failures of the exclusion restriction that invalidate the IV strategy. For instance, ethnic concentration at assignment may be correlated with availability of information on contraception or number of pharmacies which influence usage of contraceptives. I still report results from this strategy to see if they give the same conclusion than the ITT estimates which are less subject to criticisms. In specifications where ethnic concentration enters quadratically, I need two instruments. I use concentration at assignment and its square. For the control function in a duration model, I follow Coviello et al. (2015), estimate the two first stage linearly (separately), compute the residuals of both regressions and plug them in the hazard rate. As a robustness, I will run the OLS (naive and ITT) and the 2SLS on the subsample of complete and incomplete spells. I do not estimate duration models on complete and incomplete spells since I may not well measure the time of failure.

6.2 Naive Estimation

Results are summarized in table 13, for immigrants from the same region and in table A8 when concentration is measured with all non-Western immigrants. Panel A reports estimations of equation 1.4 and panel B of equation 1.5. The naive estimation is negative, doubling the size of the ethnic community (from the same region) is associated with a decrease in the probability of having taken contraceptives by the age of 20 y.o., of around 1.5 percentage points. This negative relation is similar when using a duration model, meaning that non only the final point of the adoption of the behaviour is affected but also the process leading to it. Results (5) are of the same order of magnitude than in other columns but are only marginally significant for the Weibull regressions and not significant when estimating with OLS. Part of the variation in ethnic concentration is explained by municipalities, thus including municipality fixed effects and focusing on the difference between zip4 within the same municipality reduces the variation used for estimation and increases the standard errors.

Table 13 here

The results are lower in magnitude and insignificant when one pools together all non-Western immigrants. This points towards sorting (which is likely to bias the estimates downwards) to take place among immigrants from the same origin. Another element that stands out is that the influence of the community is not quadratic. For both measures of ethnic concentration and

both linear and non-linear models, none of the results are significant. The average age in this sample is relatively small, so the coefficient on age should be compared to the descriptive results on convergence for women of a similar age. They are of a similar order of magnitude but do not appear as significant in this specification due to smaller sample size as compared to the descriptive part of the paper.

6.3 Intention to Treat Estimation

In tables 14 and A9 are reported the ITT results for the two measures of ethnic concentration. The results are very small in magnitude and not statistically significant from zero. When one compares OLS estimates from table 13 and 14, they were divided by two for specifications that include municipality fixed effects and from 6 to 15 for specifications (1) to (4). Results are non significant also for the duration analysis. With very few controls, i.e. specifications (1) and (2), estimates show a doubling of the size of the ethnic community changing the hazard rate of starting to take contraceptives by less than 0.1 percentage points. For the Weibull MLE, the sign of the estimates also change between specifications.

Table 14 here

The coefficients are smaller in magnitude than those of age at migration. The effect does not seem to be quadratic and is similar for both definitions of ethnic concentration. The fact that neighbourhood ethnic concentration does not change neither the speed at which immigrants adopt the behaviours of natives, nor the final outcome of this process is the main result from this paper.

6.4 Instrumental Variable Estimation

In tables 15 and A10 are reported the instrumental variable results. As said above, this strategy is not free of flows since it can be plagued by failure of the exclusion restriction. However, finding a similar zero effect with this approach gives additional credit to the ITT estimates. In particular, having shown evidence supporting monotonicity, if one is willing to accept that exclusion is satisfied, the IV estimate is a rescaled version of the ITT Imbens and Angrist (1994); Angrist et al. (1996); Heckman et al. (2006). The IV estimates then have the same sign as the ITT but are larger in magnitude. The estimate also identifies an effect on a subpopulation of compliers for which it will be larger than its average on all assigned to treatment.

Table 15 here

A concern throughout this paper is that the assignment is not meaningful in the sense that asylum seekers spent too little time in the area (centre and later close-by neighbourhoods) where they were assigned. Having very low ITT estimates is compatible with the two explanations,

neighbourhood having no effect and assignment not being influential because the length of exposure is too small. Looking at the IV estimates focuses on a subpopulation for which assignment was meaningful since it pushed it to comply and live in more ethnically concentrated areas. It also inflates the coefficients and potentially allows to detect an effect. Looking at the strength of the first stage also allows to test for meaningful assignment.

Results also point towards a zero result both for the 2SLS and control function specifications. This result is all the more convincing that first stage F test are large for specifications (1) to (4), around 30 which is way larger than the rule of thumb of 10. The instrument appears weak in specifications where municipality fixed effects are included. As pointed before, part of the variation in ethnic concentration at the zip4 level comes from being in a municipality with a larger/smaller population of immigrants. Including municipality FE in the first stage, regressions thus limit the relation between concentration at age 15 and at assignment since part of it is explained by municipality. Instrumenting concentration and its square at age 15 by concentration and its square at assignment provide a weak instrument, which is why I do not comment those results but show them for completeness.

An alternative IV strategy would be to instrument concentration at age 15 with the number of asylum seekers who have been assigned to the same zip4 in the years prior to assignment (as in Damm (2009)). Results of this strategy are reported in table A12. The F test from the first stage are extremely low. This alternative strategy is not possible in this setting.

7 Heterogeneity and potential mechanisms

Looking at treatment effect heterogeneity is interesting for two reasons. First, it can be that the zero result is the average between a positive and a negative result or that the result is significantly positive/negative for some meaningful subgroups. Second, looking into some dimension of heterogeneity is indicative of potential mechanisms at play.

The identification strategy is the strength of the paper since it is not easy to observe an exogenous variation in residential choices. It is also a limitation since few mechanisms can be tested within this setting. However, few does not mean none and still a few things can be done. If the results are very different depending on age at arrival, for instance are much stronger when girls are very young, this would give an indication that immigrant women form very early their conception of sexuality and can only be influenced at a young age.

I look at three particular dimensions of heterogeneity: (i) whether a young woman arrived below or above the mean age of arrival, (ii) whether she stayed in a COA accommodation for a period longer or shorter than the average of other girls and (iii) whether the centre was smaller/largerin terms of number of registered asylum seekers, than the mean size in the sample[25]. I look at

[25]The mean age of arrival is ..., other statistics can be found in table 7.

below/above the mean rather than the median mostly to emphasize on long stays in centres[26]. The mean duration of stays is higher than the median (9 vs 15 months) and I want to make sure that the result is zero even for people who stayed for a long time in the assigned neighbourhood.

Since, this part of the analysis is not reported in the pre-analysis plan, I should use specific econometric technique such as causal trees, Athey and Imbens (2016); Wager and Athey (2018), in order to account for multiple hypothesis testing when looking at heterogeneity of effects. However, I fail to reject the null in all of the cases that I investigate below. For simplicity, I thus rely on the same standard type of inference than I used in the last section. I estimate the following equation

$$y_{i,t} = \alpha + \beta_1 \ln e_{j,g,t} \mathbb{1}\{\text{below mean}\} + \beta_2 \ln e_{j,g,t} \mathbb{1}\{\text{above mean}\} + \pi X_i + \epsilon_{i,t} \tag{1.6}$$

where $\mathbb{1}\{\text{below mean}\}$ is an indicator function that takes the value one if the age at arrival, length of stay or number of asylum seekers at the centre are below the sample mean. I estimate both linear regressions and duration models, each of them as an ITT and as an instrumental variable/control function. I use the interaction of concentration at age 15 and below/above the mean as instruments.

Results are reported in tables 16 and A11, respectively for immigrants from the same origin and immigrants from all non Western origins. Panels A and B report results from linear estimations (OLS and 2SLS), while panels C and D report results of Weibull regressions as an ITT (panel C) and as a control function (panel D). For each regression, I test $\mathbb{H}_0 : \beta_1 = \beta_2$ and report the p-value of this test. I only report specifications (3) to (5) to not clutter the tables.

Table 16 here

Age at arrival and Size of the centres There is no heterogeneity between arriving below or above the mean age in the sample. Results are small in magnitude, not statistically different from zero and the effect is not statistically different below and above the mean. This result holds for both linear and non-linear estimations, ITT and IV, different definitions of peer groups (all non-Western immigrants and immigrants from the same origin). The same holds for being in a centre of smaller or larger size.

Length of stay in the centres The picture is slightly different when one looks at above/below mean duration in COA centres. The results are indeed small in magnitude and not statistically significant from zero but the effect is different for people staying more than the average time and those staying less. In 8 out 12 specifications the p-value of the test of coefficient equality is below 0.05 in table 16 and for the remaining four specifications, the p-value is below 10%. The picture is less clear cut but also holds in table A11. This evidence is however not strong enough to conclude

[26]Results using the median are similar and available upon request.

that neighbourhoods affect the convergence in cultural behaviour in one way or the other since no single coefficient is itself significant.

The fact that so many estimation techniques and subsamples give the same result show that cultural behaviour is not influenced by the environment at least not when it is accounted for at the neighbourhood level.

8 Robustness Checks

This paper establishes a null result that cultural convergence is not affected by ethnic concentration at the neighbourhood level. There are three ways in which this result may be invalidated. First, the zip4 level may not be the relevant level. It could be small and miss part of the social interactions. Second, it could be that the zero result is in fact a noisy effect that would become significant if the sample size was larger. Third, it could be that the entire strategy is not able to capture any effect for any outcome. This could be the case if for instance, asylum seekers stay too little in the centres in which they were assigned. The assignment would not be meaningful and the strategy not able to capture any effect. This concern has already been addressed, by looking at the heterogeneity between individuals who stay for longer in the COA accommodation and by using an IV strategy. However, further checks can be give additional credit to the validity of the setting.

To address the three concerns mentioned, I do the following: first I reproduce the baseline analysis, meaning the naive, ITT and IV on complete spells but instead of calculating concentration at the zip4 level, I calculate it at the municipality level. Second, I reproduce the baseline analysis on the sample of complete and incomplete spells, meaning all the women who are observed until age 20 which increases the sample size by a third. Finally, I reproduce the entire analysis (baseline and heterogeneity) with a different outcome. Instead of focusing on culture, captured by usage of contraceptives, I focus on educational outcomes. Previous work Aslund et al. (2011) has shown a positive effect of ethnic enclaves on GPA in Swedish high schools. I check whether I can capture a similar effect in the Netherlands. The objective is to determine whether the empirical strategy used in this paper can establish any statistically significant result. In a sense this is the reverse of a placebo test, I take an outcome on which previous literature has found an effect to see if I can find one in the Netherlands.

8.1 Doing the analysis at the municipality level

In tables 17, 18 and 19, I reproduce the estimation of equations 1.4 and 1.5 for the case where ethnic concentration includes other immigrants from the same origin, while tables A13, A14 and A15 do the same for all non-Western immigrants. Concentration is calculated at the municipality level rather than at the zip4 level.

A few results change: in table 17 the naive estimates are negative but much smaller in magni-

tude and non significant (only marginally for some specifications of the duration model) than they were when focusing at the zip4 level. This is probably indicative that sorting takes place at the neighbourhood level rather than at the municipality level. Thus the co-founding effect of sorting is less salient when information is collapsed at the municipality level. Besides this difference, the other results are very similar, i.e. the ITT and IV point towards a zero effect and the effect is not quadratic. This holds also when ethnic concentration is measured by all non-Western immigrants.

Tables 17, 18 and 19 here

A few additional things are worth mentioning: the ITT estimates in table 18 are now negative when those in table 14 were positive. In both cases they are anyway small in magnitude but observing a change in sign when the absolute values of the coefficients are very small probably mean that too much importance should not be given to the sign regardless of the significance. Another element that can be pointed out is that the instrument in the linear specification has much higher F-test (18.02) in specification (5), with municipality of assignment fixed effects, reinforcing the zero result that the IV estimates produce.

8.2 Increasing the sample size

A noisy estimate could be mistaken for a zero effect. This does not seem likely as the sample sizes are already quite large, almost 5,000 individuals and that standard errors would allow to estimate an effect of +/- 1.2 percentage points (for a mean outcome of 42.7%) for the ITT estimation in the linear case. However, to make sure that the effect is truly zero, one would like to estimate equations 1.4 and 1.5 with a larger sample. A way to do it is to look at the results on the sample of complete and incomplete spells. Results of the linear regressions are reported for the linear and quadratic models in tables 20, 21 and 22. The left part of the table focuses on peers from the same region while the right focuses on all non-Western immigrants.

Tables 20, 21 and 22 here

A difference is that by including the incomplete spells, one includes women who were older than 15 in 2006 and thus could for whom a complete spell cannot be observed. As a result this sample is older, the average age switch from 6.8 to 7.8 years old. The picture remains dramatically the same. The naive estimates are negative and significant. The coefficients of the left part of table 20 are almost identical to those of table 13. The estimates in column (5) are now significant. Effects of non-Western immigrants are also now significant through a combination of increase in the magnitude of the effect and and decrease in the sample size. One noticeable element is that the coefficient on age at arrival is now significant. This is mostly due to an increase in the magnitude of the results. Recall that this sample arrived older in the Netherlands and recall from table 4 that the convergence is weaker for people who arrive very young in the country.

When one turns to the causal estimates, the results are almost unchanged with none of the coefficients being even marginally significant. These results stand in contrast with the coefficient on age at migration being highly significant in all cases. The effect is not quadratic either. This is definite evidence that the estimates here are a well identified zero and not a noisy effect.

8.3 Falsification test: Looking at another outcome

A concern of the paper is that the identification strategy is not able to capture any effect of any sort, on any outcome. This could be the case if for instance assignment to a centre is not a meaningful residential anchor. Although this concern has primarily been addressed in previous sections, I add an additional robustness check. I take an outcomes on which previous research has shown that there was an effect and I see whether this strategy can capture it. In a sense, I do a falsification test but the other way around. Instead of looking at a placebo, I focus on something that works to establish that a strategy can detect an effect. In return, this gives more credit when regressions produce a zero result.

Because this part of the analysis is not pre-registered, I want to explain my choice of outcome variable. Previous research Aslund et al. (2011) has shown the effect of living in an ethnic enclave on educational achievements of young immigrants. I want to choose an outcome as close as possible to theirs to minimize the risk of manipulation. I also want the falsification test to be as close as possible to the main analysis of this paper. In particular, I want the outcome to be a binary variable, ideally with the same sample mean. Therefore, I focus on an outcome has completed such or such level of education by age 20. I focus on the same sample, meaning women with complete and incomplete spells. In the falsification test, there is no ex-ante reason to exclude women whose early teenage years are not observed. However, to mimic the analysis done in this paper, one should stick to the sample of complete spells. I report results from both samples.

The analysis by Aslund et al. (2011) focused on GPA in high school. I look at the outcome has followed an education level (in the Netherlands) as high as the level (in Sweden) for which Aslund et al. (2011) use GPA[27]. Results from the estimation of equations 1.4 and 1.5 reported in tables 23, 24 and 25 respectively for the naive, ITT and IV specifications. The left of the tables report the estimates from using the complete spells while the right part of the table corresponds to estimates when using the complete and incomplete spells. Similar tables for all immigrants can be found in tables A16, A17 and A18.

Tables 23, 24 and 25 here

The naive estimates are insignificant for all specifications and measures of ethnic concentration

[27]To identify the corresponding level in the Netherlands, I rely on an online document that can be found at the adress https://www.nuffic.nl/en/publications/education-system-sweden/ in particular on page 5. Compuslory education in Sweden corresponds to at least HAVO, VWO in the Netherlands. I focus on the highest education followed not completed to account for those who did not pass but would have a GPA in the Swedish example.

but for the effect on all non-Western immigrants in the case where municipality fixed effects are included. In the larger sample, the estimates are lower in magnitude than in the smaller ones and in all cases not significantly different from zero. The coefficients of age at migration change sign between the smaller and larger sample and has a negative sign when estimated on complete and incomplete spells.

The ITT estimates are also not statistically different from zero for most specifications but are at least as large in magnitude as the naive estimators with the opposite sign with complete and incomplete spells, the specification that controls for neighbourhood quality, the *status score*, is marginally significant. The situation is similar for IV estimates. There is no evidence of a quadratic relationship.

In addition to this baseline analysis, I estimate equation 1.6 to uncover potential sources of heterogeneity. Results are reported in table 26 and A19. The two upper panels, panel A and B report the estimates of the ITT and IV strategies for the complete spells and the two lower ones for the complete and incomplete spells. There is evidence that the effect is particularly negative for those staying in a smaller centre. This could be the case if asylum seekers primary network is made of people from their centre. Staying in a smaller one means having fewer people to interact with. This can be detrimental for their studies.

Table 26 here

The point of this robustness check is not to determine the causal effect of living in an ethnic enclave on educational achievements. The estimates point towards a negative result but the effects are not always clear cut. However, in similar conditions (binary variable, with larger mean but similar variance, same samples) but with an educational outcome, one does not observe that the ITT or the IV shrink to zero with respect to the OLS naive estimate. As the sample size grows, the naive estimator gets closer to zero and is not significant while the ITT and IV become marginally significant and relatively large for some subgroups. The picture is different than the uniformly zero effect of ethnic concentration on usage of contraceptives.

This robustness check together with the descriptive evidence gives credibility to the causal estimates of neighbourhoods. The descriptive evidence established that the outcome could change and this last robustness that neighbourhoods could affect another outcome than usage of contraceptives. If neither the outcome, nor the policy only produce zero results, one can be confident when the estimates point to one.

9 Conclusion

In this paper, I study the cultural behaviour of immigrants, see if it converges to that of natives and if this process is influenced by neighbourhood ethnic composition. I measure culture with

the usage of contraceptives by young women. Taking contraceptives, while a very personal and intimate choice, is available in administrative registries. In a first round of descriptive analysis, I show that there are very strong differences in contraceptive usage between immigrants and natives but that there is evidence of convergence in behaviours: immigrants who arrived older are less likely to take them. I then exploit the placement of asylum seekers in temporary welcome centers as an experiment which brings exogenous variation in neighbourhood characteristics. I find that growing up in a environment with a large ethnic community has no effect on the probability of using contraceptives in the ages 15 to 20. This empirical result is very robust, i.e. holds for various definitions of ethnic concentration and with several estimations strategies.

By studying neighbourhoods, I pool together people from different ages who may have weaker/stronger ties to young immigrants. Future research should focus on the potential influence of different peer groups from different social activities who are potentially closer to the population of interest. Since this paper relies on an exogenous variation in residential choices, it is beyond the scope of this work to test the role of these alternative interactions for which alternative identification strategies must be found. This paper reaches, however, a clear conclusion on a question of general interest: ethnic enclaves do not slow down the adoption of natives cultural behaviours by immigrants.

Bibliography

Algan, Y., Mayer, T., and Thoenig, M. (2013). The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France. CEPR Discussion Papers 9416, C.E.P.R. Discussion Papers.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Aslund, O., Edin, P.-A., Fredriksson, P., and Groenqvist, H. (2011). Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy. *American Economic Journal: Applied Economics*, 3(2):67–95.

Aslund, O. and Fredriksson, P. (2009). Peer Effects in Welfare Dependence: Quasi-Experimental Evidence. *Journal of Human Resources*, 44(3).

Aslund, O., Fredriksson, P., and Edin, P.-A. (2003). Ethnic Enclaves and the Economic Success of Immigrants?Evidence from a Natural Experiment. *The Quarterly Journal of Economics*, 118(1):329–357.

Aslund, O. and Rooth, D.-O. (2007). Do when and where matter? initial labour market conditions and immigrant earnings. *Economic Journal*, 117(518):422–448.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Science*, 113(27):7353–7360.

Bailey, M. J. (2006). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply. *The Quarterly Journal of Economics*, 121(1):289–320.

Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *Review of Economic Studies*, 79(1):128–161.

Beckers, P. and Borghans, L. (2011). Segregation in neighbourhoods and labour market outcomes of immigrants: Evidence from random assignment in the Netherlands. Technical report.

Bertrand, M. and Kamenica, E. (2018). Coming Apart? Cultural Distances in the United States over Time. NBER Working Papers 24771, National Bureau of Economic Research, Inc.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network Effects and Welfare Cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2008). Are Muslim Immigrants Different in Terms of Cultural Integration? *Journal of the European Economic Association*, 6(2-3):445–456.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Errata Corrige: Are Muslim Immigrants Different In Terms Of Cultural Integration? *Journal of the European Economic Association*, 9(5):1012–1019.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like Beckham: Ethnic identity and integration. *European Economic Review*, 90(C):146–164.

Bisin, A., Topa, G., and Verdier, T. (2004). Religious Intermarriage and Socialization in the United States. *Journal of Political Economy*, 112(3):615–664.

Bisin, A. and Verdier, T. (2000). Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits. *Quarterly Journal of Economics*, 115(3):955–988.

Bisin, A. and Verdier, T. (2001). The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory*, 97(2):298–319.

Bisin, A. and Verdier, T. (2010). The Economics of Cultural Transmission and Socialization. In Jess Ben-habib, A. B. and Jackson, M., editors, *Handbook of Social Economics*, volume 1A of *Handbook of Social Economics*, chapter 9, pages 339–416. Elsevier.

Chetty, R. and Hendren, N. (2018a). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R. and Hendren, N. (2018b). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4):855–902.

Coviello, D., Ichino, A., and Persico, N. (2015). The Inefficiency Of Worker Time Use. *Journal of the European Economic Association*, 13(5):906–947.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045.

Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, 27(2):281–314.

Damm, A. P. and Dustmann, C. (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review*, 104(6):1806–32.

Danzer, A. M., Feuerbaum, C., Piopiunik, M., and Woessmann, L. (2018). Growing Up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children. IZA Discussion Papers 11608, Institute of Labor Economics (IZA).

Fernandez, R. (2013). Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *American Economic Review*, 103(1):472–50.

Fogli, A. and Veldkamp, L. (2011). Nature or Nurture? Learning and the Geography of Female Labor Force Participation. *Econometrica*, 79(4):1103–1138.

Goldin, C. and Katz, L. F. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, 110(4):730–770.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119.

Kohler, H.-P., Behrman, J. B., and Watkins, S. C. (2001). The Density of Social Networks and Ferility Decisions: Evidence from South Nyanza District, Kenya. *Demography*, 38(1):43–58.

Kuran, T. and Sandholm, W. (2008). Cultural Integration and Its Discontents. *Review of Economic Studies*, 75(1):201–228.

Lalonde, R. and Giguère, B. (2008). When might the two cultural worlds of second generation biculturals collide? *Canadian Diversity*, 6(2):58–62.

Loken, K. V., Mogstad, M., and Wiswall, M. (2012). What Linear Estimators Miss: The Effects of Family Income on Child Outcomes. *American Economic Journal: Applied Economics*, 4(2):1–35.

Munshi, K. and Myaux, J. (2006). Social norms and the fertility transition. *Journal of Development Economics*, 80(1):1–38.

Olcina, G., Panebianco, F., and Zenou, Y. (2017). Conformism, Social Norms and the Dynamics of Assimilation. CEPR Discussion Papers 12166, C.E.P.R. Discussion Papers.

Patacchini, E. and Zenou, Y. (2016). Social networks and parental behavior in the intergenerational transmission of religion. *Quantitative Economics*, 7(3):969–995.

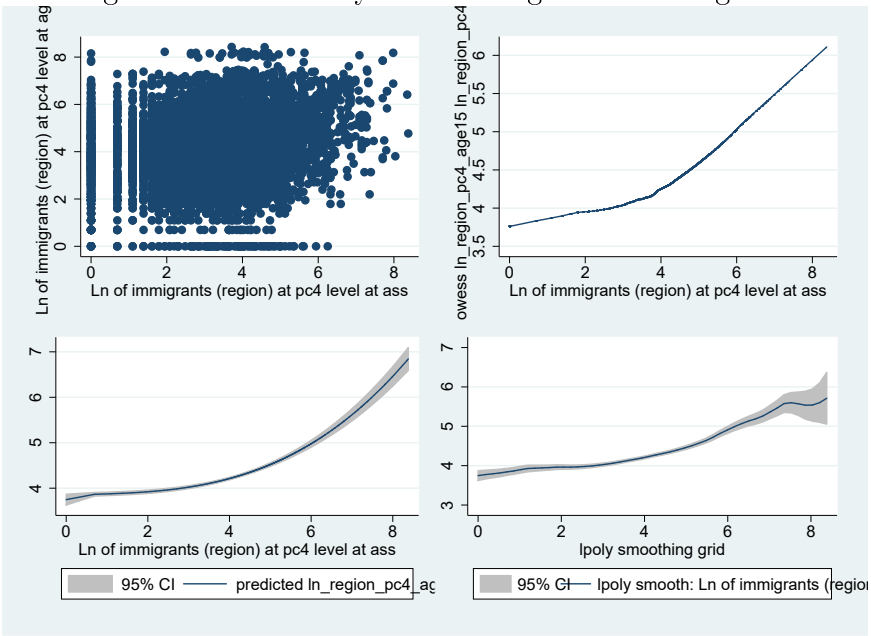
Tabellini, M. Gifts of the immigrants, woes of the natives: Lessons from the age of mas migration. *Review of Economic Studies*. forthcoming.

Verdier, T. and Zenou, Y. (2017). The role of social networks in cultural assimilation. *Journal of Urban Economics*, 97(C):15–39.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Figure 1: Community size at assignment and age 15



Note : I plot the log of ethnic concentration at age 15 against the log of ethnic concentration at the time of assignment. Ethnic concentration is defined as immigrants from the same region and neighbourhood refers to zip4. The figure is split into four quadrants, in the upper left one, I scatter the points, in the upper right one I approximate the scatter of points with a locally weighted regressions, in the lower left one I use a quadratic approximation and in the lower right one a kernel-weighted local polynomial.

Table 1: Difference in usage between natives and immigrant women

Age	Nb of observations			% of contraceptives use		
	Natives	1st gen	2nd gen	Natives	1st gen	2nd gen
15	524098	19344	102609	24	7	11
16	523777	19215	102352	47	15	23
17	523573	19187	102233	65	26	34
18	523391	19152	102123	77	35	44
19	523335	19168	102120	83	44	51
20	524098	19344	102609	87	50	57

Note : This table reports the number of girls who were living in the Netherlands between the ages of 15 to 20 and the percentage who have used contraceptives at least once by a certain age. Girls are classified into three groups, natives, first and second generation immigrants from non-western countries. Columns 2 to 4 report the total number of girls while the last three columns report the percentage of usage of contraceptives. Elements (4,2) and (4,5) off the table should be read as follows: 523,391 native girls were 19 y.o. between 2006 and 2017 and 77% of them had used contraceptives at least once at age 19.

Table 2: Difference in usage between first and second generation

Age	Nb Obs		% of contra use	
	1st	2nd	1st	2nd
Immigrants from Turkey				
15	1249	21680	3	4
16	1246	21619	7	9
17	1245	21578	11	15
18	1240	21551	16	21
19	1238	21584	23	28
20	1249	21680	32	34
Immigrants from Morocco				
15	1114	19409	2	3
16	1112	19386	5	7
17	1111	19364	9	12
18	1108	19333	13	17
19	1106	19284	20	24
20	1114	19409	27	30
Immigrants from Surinam				
15	1085	15687	9	14
16	1079	15651	21	29
17	1082	15638	34	44
18	1076	15613	48	58
19	1078	15608	58	66
20	1085	15687	67	73
Immigrants from Indonesia				
15	180	4807	6	20
16	179	4804	18	39
17	178	4799	33	57
18	177	4797	50	70
19	179	4796	60	77
20	180	4807	70	82
Immigrants from Yugoslavia				
15	432	2257	5	13
16	431	2249	12	28
17	430	2246	22	42
18	427	2246	30	55
19	429	2245	42	63
20	432	2257	53	70

Note : This table reports the number of first and second generation immigrant girls from five different countries who were living in the Netherlands between the ages of 15 to 20. Columns 2 to 4 report the total number of girls while the last two columns report the percentage who has used contraceptives at least once by a certain age.

Table 3: Descriptive Analysis - Baseline Results

Age	(A)	(B)	(C)	Mean	Nb obs
		Panel A: Linear regressions			
15	-0.0051***	-0.0072***	-0.0073***	6.6	19394
	(0.00047)	(0.00097)	(0.001)		
16	-0.0076***	-0.0093***	-0.0091***	15.2	19257
	(0.00067)	(0.00138)	(0.00143)		
17	-0.0087***	-0.0105***	-0.0103***	25.4	19228
	(0.0008)	(0.00164)	(0.0017)		
18	-0.0095***	-0.011***	-0.0109***	35	19193
	(0.00086)	(0.00178)	(0.00184)		
19	-0.0085***	-0.009***	-0.0094***	43.2	19205
	(0.00089)	(0.00184)	(0.00191)		
20	-0.0087***	-0.0072***	-0.0075***	50.1	19394
	(0.0009)	(0.00185)	(0.00193)		
		Panel B: Duration models			
	-0.026***	-0.028***	-0.029***	34	98423
	(0.0015)	(0.0031)	(0.0032)		
Country FE	✓	✓	✗		
Year of arrival FE	✗	✓	✗		
Country / year FE	✗	✗	✓		

Note : Each cell in the panel A reports the results of a linear regression of the dummy having taken contraceptives at least once by a certain age on age at arrival and a series of controls. The coefficient reported is that of age at arrival, the first column reports results with country fixed effects, the second with (separate) country and year fixed effects and the last one with fixed effects for the interaction of country and year of arrival. Panel B reports the results of a Weibull regression of the same specifications. The regressions are estimated on the sample of first generation women who were living in the country between 15 and 20 during the period 2006 and 2017.

Table 4: Descriptive Analysis - Heterogeneity by Age at arrival

Age	$\phi(\lambda_c, \theta_t) = \lambda_c$			$\phi(\lambda_c, \theta_t) = \lambda_c + \theta_t$			$\phi(\lambda_c, \theta_t) = \pi_{c,t}$			Mean	Nb obs
	0-5 y.o.	6-10 y.o.	11-15 y.o.	0-5 y.o.	6-10 y.o.	11-15 y.o.	0-5 y.o.	6-10 y.o.	11-15 y.o		
Panel A: Linear regressions											
15	-0.0065***	-0.0047***	-0.005***	-0.0082***	-0.0072***	-0.0069***	-0.0073***	-0.0068***	-0.0077***	6.8	18770
	(0.0022)	(0.001)	(0.0007)	(0.0023)	(0.0013)	(0.0011)	(0.0024)	(0.0014)	(0.0011)		
16	-0.0092***	-0.0063***	-0.0076***	-0.0107***	-0.0085***	-0.0092***	-0.0098***	-0.0077***	-0.009***	15.4	18656
	(0.0031)	(0.0015)	(0.0009)	(0.0033)	(0.0019)	(0.0015)	(0.0034)	(0.0019)	(0.0016)		
17	-0.0069	-0.0061***	-0.0077***	-0.0089**	-0.0087***	-0.0094***	-0.0081*	-0.0082***	-0.0092***	25.7	18629
	(0.0037)	(0.0017)	(0.0011)	(0.0039)	(0.0022)	(0.0018)	(0.0041)	(0.0023)	(0.0019)		
18	-0.0109***	-0.0077***	-0.0092***	-0.0123***	-0.0097***	-0.0107***	-0.0104**	-0.0088**	-0.0103***	35.3	18590
	(0.004)	(0.0019)	(0.0012)	(0.0042)	(0.0024)	(0.0019)	(0.0044)	(0.0025)	(0.002)		
19	-0.0117***	-0.0086***	-0.0086***	-0.0122**	-0.0097***	-0.0088***	-0.0118***	-0.0097***	-0.0089***	43.4	18595
	(0.0041)	(0.0019)	(0.0012)	(0.0044)	(0.0024)	(0.002)	(0.0046)	(0.0026)	(0.0021)		
20	-0.0059	-0.006***	-0.0076***	-0.0043	-0.0049**	-0.0061***	-0.003	-0.0047*	-0.0061***	50.3	18770
	(0.0042)	(0.0019)	(0.0012)	(0.0044)	(0.0025)	(0.002)	(0.0046)	(0.0026)	(0.0021)		
Panel B: Duration models											
	-0.0269***	-0.0216***	-0.244***	-0.0289***	-0.0255***	-0.0262***				34.2	95370
	(0.0068)	(0.0032)	(0.002)	(0.0073)	(0.0041)	(0.0033)					
Country FE	✓	✓	✓	✓	✓	✓	✗	✗	✗		
Year of arrival FE	✗	✗	✗	✓	✓	✓	✗	✗	✗		
Country / year FE	✗	✗	✗	✗	✗	✗	✓	✓	✓		

Note : Panel A reports the results of a linear regression of the dummy having taken contraceptives at least once by a certain age. The explanatory variables are age at arrival interacted with a dummy for arriving before age 5, between 6 and 10 and between 11 and 15 and a series of controls. The coefficient reported are those of age at arrival interacted with the age group dummies. The first three columns report results with country fixed effects, the following three with (separate) country and year fixed effects and the last three with fixed effects for the interaction of country and year of arrival. Panel B reports the results of a Weibull regression of the same specifications. The regressions are estimated on the sample of first generation women who were living in the country between 15 and 20 during the period 2006 and 2017.

Table 5: Descriptive Analysis - Heterogeneity by Country of Origin

Age	Turkey	Morocco	Surinam	Afghanistan	Iraq	Mean	Nb Obs
Panel A : Country of Origin							
15	-0.0003	-0.0015	-0.0046***	-0.0009	-0.0029***	4.3	7981
	(0.0013)	(0.0015)	(0.0017)	(0.0013)	(0.0012)		
16	.0003	-0.0016	-0.0083***	-0.0031	-0.0053***	9.9	7940
	(0.0019)	(0.0021)	(0.0024)	(0.0019)	(0.0018)		
17	.0003	-0.0012	-0.0081***	-0.0051**	-0.0072***	16.7	7923
	(0.0023)	(0.0026)	(0.003)	(0.0023)	(0.0022)		
18	-0.0024	-0.0022	-0.0049	-0.0069***	-0.0104***	23.7	7906
	(0.0026)	(0.003)	(0.0034)	(0.0026)	(0.0025)		
19	-0.0023	-0.0037	-0.0026	-0.0089***	-0.0113***	31.2	7915
	.(00028)	(0.0032)	(0.0037)	(0.0028)	(0.0027)		
20	-0.0023	-0.0033	-0.0057	-0.0095***	-0.0123***	38.4	7981
	(0.003)	(0.0034)	(0.0039)	(0.003)	(0.0029)		
Panel B : Country of Origin and Year of Arrival							
15	-0.0027	-0.0036**	-0.0073***	-0.0025	-0.0058***	4.3	7981
	(0.0017)	(0.0017)	(0.002)	(0.0016)	(0.0017)		
16	-0.0022	-0.0031	-0.0111***	-0.0045*	-0.0081***	9.9	7940
	(0.0025)	(0.0026)	(0.003)	(0.0024)	(0.0025)		
17	-0.0028	-0.0034	-0.0115***	-0.0068***	-0.0104***	16.7	7923
	(0.0031)	(0.0032)	(0.0036)	(0.0029)	(0.003)		
18	-0.005	-0.0043	-0.0081**	-0.0087***	-0.0132***	23.7	7906
	(0.0035)	(0.0036)	(0.0041)	(0.0033)	(0.0034)		
19	-0.0048	-0.0057	-0.006	-0.0113***	-0.0143***	31.2	7915
	(0.0038)	(0.0039)	(0.0045)	(0.0036)	(0.0037)		
20	0.0004	-0.0011	-0.0044	-0.0076**	-0.0106***	38.4	7981
	(0.004)	(0.0041)	(0.0047)	(0.0037)	(0.0039)		
Panel C : Country of origin interacted with year of arrival							
15	-0.0043**	-0.0021	-0.0056	-0.0048*	-0.0025	4.3	7981
	(0.0022)	(0.003)	(0.0033)	(0.0026)	(0.003)		
16	-0.0054*	-0.0047	-0.0036	-0.0115***	.0005	9.9	7940
	(0.0032)	(0.0044)	(0.0049)	(0.0039)	(0.0043)		
17	-0.0043	-0.0035	-0.0069	-0.0173***	-0.0017	16.7	7923
	(0.0039)	(0.0054)	(0.006)	(0.0048)	(0.0053)		
18	-0.0077*	-0.0039	-0.0026	-0.0208***	-0.0004	23.7	7906
	(0.0045)	(0.0062)	(0.0068)	(0.0054)	(0.006)		
19	-0.008	-0.0047	-0.003	-0.0229***	-0.0014	31.2	7915
	(0.0048)	(0.0067)	(0.0074)	(0.0059)	(0.0066)		
20	-0.003	0.0042***	-0.0052	-0.0224***	0.0046	38.4	7981
	(0.0051)	(0.007)	(0.0077)	(0.0061)	(0.0069)		

Note : Each row reports the results of a linear regression of the dummy having taken contraceptives at least once by a certain age. The explanatory variables are age at arrival interacted with a dummy for five countries of origin (Iraq, Afghanistan, Turkey, Morocco and Surinam) and a series of controls. The coefficients reported are those of age at arrival interacted with the country dummies. The first five rows report results with country of origin fixed effects, the following with (separate) country of origin and year of arrival fixed effects and the last five with fixed effects for the interaction of country and year of arrival. The regressions are estimated on the sample of first generation women from Iraq, Afghanistan, Turkey, Morocco and Surinam who were living in the Netherlands between 15 and 20 during the period 2006 and 2017.

Table 6: Summary Information on COA accommodation

	Nb of centre appearing at the same address				Brief description
Type of centre	1	2	3	4	
Aanvullende opvang	183	45	13	NA	Collective housing, for emergency when no other location available
Administratief geplaatst	2,354	859	199	42	Individual housing found by a.s. with contacts in the NL
Alternatieve tijdelijke capaciteit	15	NA	NA		
Asielzoekerscentrum	66	88	45	23	Collective housing
Contingent	52	31	17	NA	
Gemeentewoning	3,698	442	83	21	Individual housing for a.s. after being granted refugee status
Kinderwoongroep	20	NA	NA	NA	Small scale location for unaccompanied minors
Kleinschalige Centrale Opvangeenhdn	1,523	182	57	26	Small scale location for unaccompanied minors
Kleinschalige wooneenheid	15	NA	NA		Small scale location for unaccompanied minors
Opvang en Onderzoekcentrum	17	NA	NA	NA	Ter Apel centre where all a.s. start the application process
Orientatie & Inburgeringslocatie	16	24	12		Centre for people being denied refugee status prior to leaving
Terugkeerlocatie	61	37	23	NA	Centre for families being denied refugee status prior to leaving
Tijdelijke Noodvoorziening	23	21	15	NA	Collective housing, for emergency when no other location available
Zelf Zorg Arrangement	7,375	2,362	442	96	Individual housing found by a.s. with contacts in the NL
Total	15,321	4,170	966	288	

Note : The first column lists the main types of accommodations while the last column briefly describes them. The second column shows the distribution among types of centres for addresses that identify a single centre. The third column shows the distribution among types of centres for addresses that identify two centres. The fourth column shows the distribution among types of centres for addresses that identify three centres. The fifth column shows the distribution among types of centres for addresses that identify four centres. The entry (2,4), i.e. 199 should be read as follows: among the addresses under which three centres are listed, 199 of them are *Administratief geplaatst*, meaning the same 199 addresses appear again twice in the same column under different types of centres. NA refers to entries where the number is below 10.

Table 7: Various information on COA centres

	Complete Spells		Complete and Incomplete Spells	
	Length of stay	Nb of other a.s.	Length of stay	Nb of other a.s.
10th p	44	0	44	0
25th p	97	23	97	17
Median	253	112	256	86
75th p	656	211	650	187
90th p	1096	334	1067	301
Mean	468	140	459	124
Nb Obs	4909	5384	7851	8661

Note : Columns 2 and 4 report the main quantiles and the mean of the distribution of length of stay in COA centres assigned at arrival. Columns 3 and 5 report the same statistics for the number of asylum seekers registered in COA centres the 1^{st} of January of the year of assignment. The population in columns 2 and 3 are assigned women who were in the Netherlands between ages 15 and 20 and whose usage of contraceptives could be observed for all ages from 15 to 20. The population in columns 4 and 5 are assigned women who were in the Netherlands between ages 15 and 20 and whose usage of contraceptives could be observed only partially for ages 15 to 20.

Table 8: Mobility of the experimental population

	In zip 4			Adjacent zip 4			In municipality		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
After 1 year	4886	2083	42.6	4886	2210	45.2	4886	2314	47.4
After 2 years	4849	1213	25	4849	1401	28.9	4849	1543	31.8
After 3 years	4837	628	13	4837	854	17.7	4837	1035	21.4
After 4 years	4822	366	7.6	4822	591	12.3	4822	778	16.1
After 5 years	4829	275	5.7	4829	507	10.5	4829	705	14.6
After 6 years	4841	233	4.8	4841	469	9.7	4841	674	13.9
After 7 years	4811	191	4	4811	442	9.2	4811	632	13.1
After 8 years	4739	164	3.5	4739	417	8.8	4739	595	12.6

Note : For up to 8 years after arrival and for three different geographical areas (zip4, adjacent zip4 and municipality), this table reports the number of people still living in the Netherlands (1), living in the same geographical area (2) and the ratio of the two (3). Row 6, middle table should be read as follows: after 6 years in the Netherlands, 4841 women (of the experimental population with complete spells) were still living in the Netherlands, 469 in an adjacent zip4 to the one they were assigned, which represents 9.7% of that population.

Table 9: Variation in Community sizes

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	4	13	8	24	72	188	593	878
25th p	14	30	46	88	172	463	993	1742
Median	37	68	119	292	406	941	2110	5299
75th p	77	146	443	1258	838	1907	8983	18243
90th p	154	333	1760	4524	1528	3279	23637	41918
Mean	79	156	1160	2126	735	1483	15506	27046
Nb Obs	8661	8661	8661	8661	8639	8661	7961	8082

Note : This table reports the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table 10: Balancing tests at the zip4 level

	Concentration from same origin				Concentration of all non Western immigrants			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Male	-0.152***	-0.126***	0.016***	-0.029	-0.165***	-0.151***	0.029	-0.004
	(0.05)	(0.048)	(0.036)	(0.034)	(0.048)	(0.047)	(0.026)	(0.027)
Nb of children	-0.048**	-0.015	-0.024*	0.026*	-0.043**	-0.063***	-0.007	-0.004
	(0.02)	(0.018)	(0.014)	(0.015)	(0.018)	(0.019)	(0.01)	(0.012)
Age	††	0.008**	-0.001	0.006**	0.001	0.003	-0.002	0.001
	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
Basic education	-0.029	-0.018	-0.004	0.053	0.046	0.012	0.049	0.059
	(0.066)	(0.068)	(0.051)	(0.051)	(0.06)	(0.069)	(0.033)	(0.04)
Primary education	0.003	0.044	0.037	0.123**	-0.02	0.037	0.022	0.084**
	(0.068)	(0.068)	(0.053)	(0.054)	(0.065)	(0.07)	(0.037)	(0.043)
Middle school	-0.031	-0.028	0.024	0.087*	-0.071	0.018	-0.026	0.101**
	(0.062)	(0.064)	(0.05)	(0.05)	(0.063)	(0.065)	(0.037)	(0.041)
High school	0.003	-0.086	-0.003	-0.007	-0.006	-0.038	-0.029	0.028
	(0.08)	(0.083)	(0.063)	(0.062)	(0.08)	(0.082)	(0.05)	(0.052)
Higher education	-0.011	-0.26***	0.04	-0.103*	-0.016	-0.162**	0.029	-0.051
	(0.086)	(0.077)	(0.065)	(0.063)	(0.086)	(0.08)	(0.049)	(0.055)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Arrival year FE	✓	✓	✓	✓	✓	✓	✓	✓
Municipality FE (assignment)	✗	✗	✓	✗	✗	✓	✗	✗
Municipality FE (age 15)	✗	✗	✗	✓	✗	✗	✗	✓
Nb of obs	4176	3607	4176	3607	4169	3607	4169	3607
F test	0.13	3.71	0.24	3.59	0.91	1.55	1.72	2.93
p-value	0.98	††	0.95	††	0.48	0.17	0.13	0.01

Note : This table estimates equation 1.3 on the sample of household heads of women from the experimental population (with complete spells). The left part of the table measures ethnic concentration as the log of immigrants from the same region of the world in the zip4, while the right part of the table uses the log of all non Western immigrants in the zip4. In columns (1) and (3), concentration is measured the year of arrival to a COA centre, in columns (2) and (4), it is measured the year the oldest daughter turns 15. Explanatory variables include country of origin and year of arrival fixed effects together with gender of the head, age of the household head and number of children below 18 (at assignment in (1) and (3), when the oldest daughter turns 15 in (2) and (4)) and dummies for education levels, where the baseline category is missing observation (15% of the sample). Standard errors are clustered at the country of origin, municipality level (municipality at assignment for (1) and (3), at age 15 for (2) and (4)).††is used when rounded estimates are < 0.00. In columns (2) and (4), the sample is restricted to household heads who do not live in the same zip4 (when the oldest daughter turns 15) where they were assigned.

Table 11: Balancing tests at the municipality level

	Concentration from same origin				Concentration of all non Western immigrants			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Male	-0.147*	-0.299***	-0.004	-0.008	-0.198***	-0.297***	-0.003	0.008***
	(0.086)	(0.089)	(0.021)	(0.019)	(0.061)	(0.074)	(0.004)	(0.003)
Nb of children	-0.027	-0.106***	-0.017**	-0.005	-0.025	-0.107***	-0.001	-0.011***
	(0.03)	(0.032)	(0.008)	(0.009)	(0.023)	(0.028)	(0.001)	(0.002)
Age	0.007	0.008	-0.001	0.003	0.006	0.001	ŧŧ	ŧŧ
	(0.005)	(0.006)	(0.002)	(0.002)	(0.004)	(0.005)	(0.00)	(0.00)
Basic education	0.083	-0.128	-0.014	-0.01	0.032	-0.011	-0.003	0.003
	(0.109)	(0.128)	(0.029)	(0.03)	(0.086)	(0.102)	(0.005)	(0.004)
Primary education	-0.015	-0.138	0.028	0.039	0.007	-0.043	0.008*	ŧŧ
	(0.107)	(0.124)	(0.032)	(0.03)	(0.088)	(0.097)	(0.005)	(0.004)
Middle school	-0.039	-0.254**	0.021	0.019	-0.088	-0.19**	-0.003	0.006
	(0.106)	(0.117)	(0.03)	(0.029)	(0.088)	(0.095)	(0.006)	(0.004)
High school	-0.154	-0.216	-0.015	-0.031	0.021	-0.11	-0.002	0.009*
	(0.139)	(0.144)	(0.036)	(0.033)	(0.104)	(0.113)	(0.006)	(0.005)
Higher education	-0.16	-0.347**	-0.016	0.012	-0.085	-0.247**	-0.003	-0.001
	(0.14)	(0.136)	(0.039)	(0.034)	(0.119)	(0.11)	(0.008)	(0.005)
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Arrival year FE	✓	✓	✓	✓	✓	✓	✓	✓
Municipality FE (assignment)	✗	✗	✓	✗	✗	✓	✗	✗
Municipality FE (age 15)	✗	✗	✗	✓	✗	✗	✗	✓
Nb of obs	4176	3607	4176	3607	3822	3434	3822	3434
F test	1.36	1.64	0.85	1.49	0.76	2.12	1.51	1.18
p-value	0.23	0.15	0.51	0.19	0.58	0.06	0.18	0.32

Note : This table estimates equation 1.3 on the sample of household heads of women from the experimental population (with complete spells). The left part of the table measures ethnic concentration as the log of immigrants from the same region of the world in the municipality, while the right part of the table uses the log of all non Western immigrants in the municipality. In columns (1) and (3), concentration is measured the year of arrival to a COA centre, in columns (2) and (4), it is measured the year the oldest daughter turns 15. Explanatory variables include country of origin and year of arrival fixed effects together with gender of the head, age of the household head and number of children below 18 (at assignment in (1) and (3), when the oldest daughter turns 15 in (2) and (4)) and dummies for education levels, where the baseline category is missing observation (15% of the sample). Standard errors are clustered at the country of origin, municipality level (municipality at assignment for (1) and (3), at age 15 for (2) and (4)).ŧŧis used when rounded estimates are < 0.00. In columns (2) and (4), the sample is restricted to household heads who do not live in the same zip4 (when the oldest daughter turns 15) where they were assigned.

Table 12: Grouping countries into regions

Region	Countries
North Africa	Algeria Egypt Libya Morocco Sudan Tunisia
East Africa	Burundi Comoros Djibouti Eritrea Ethiopia Kenya Madagascar Malawi Mozambique Rwanda Seychelles Somalia Uganda Tanzania Zambia Zimbabwe
Central Africa	Angola Cameroon Central African Republic Chad D.R. Congo Zaire Equatorial Guinea Gabon
South Africa	Botswana Namibia South-Africa
West Africa	Benin Burkina Faso Cape Verde Ivory Cost Gambia Ghana Guinea Guinea-Bissau Liberia Mali Mauritania Niger Nigeria Senegal Sierra Leone Togo
Latin America	Antigua Barbuda Argentina Bolivia Brazil Chili Colombia Cuba Dominican Republic Ecuador El Salvador Guatemala Guyana Haiti Honduras Jamaica Mexico Nicaragua Panama Peru Surinam Trinidad Tobago Uruguay Venezuela
North America	USA Canada
Central Asia	Kazakhstan Kyrgyzstan Turkmenistan Tajikistan Uzbekistan
East Asia	China Korea South-Korea North-Korea Japan Mongolia Taiwan
South-East Asia	Cambodia Indonesia Laos Malaysia Myanmar Philippines Singapore Thailand Vietnam
South Asia	Afghanistan Bangladesh Bhutan India Iran Nepal Pakis Sri Lanka Tibet
Middle East	Armenia Azerbaijan Bahrain Cyprus Georgia Iraq Israel Jordan Kuwait Lebanon Oman Qatar Saudi Arabia Palestine Syria Turkey United Arab Emirates Yemen
Eastern Europe	Belarus Bulgaria Czech Republic Hungary Poland Moldavia Romania Russian Federation Slovakia Ukraine
Northern Europe	Estonia Latvia Lithuania Great-Britain
Southern Europe	Albania Bosnia Herzegovina Croatia Greece Italy Montenegro Portugal Slovenia Spain Yugoslavia Kosovo Macedonia
Western Europe	Austria Belgium France Germany Netherlands Switzerland
Oceania	Australia Samoa

Table 13: Baseline Results - Naive Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
					Panel A: Linear specification					
Ethnic concentration	-0.013**	-0.014**	-0.014**	-0.015**	-0.012	-0.054**	-0.055**	-0.059**	-0.065**	-0.062*
	(0.0066)	(0.0066)	(0.0066)	(0.0074)	(0.0094)	(0.0228)	(0.0229)	(0.0229)	(0.0253)	(0.0342)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.006	-0.003
		(0.0038)	(0.0039)	(0.0040)	(0.0042)		(0.0128)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.056	0.056	0.061	0.061	0.145					
					Panel B: Quadratic specification					
Ethnic concentration	-0.032	-0.032	-0.028	-0.023	-0.021	-0.174*	-0.176*	-0.174*	-0.155	-0.186
	(0.0286)	(0.0286)	(0.0285)	(0.0293)	(0.0330)	(0.1018)	(0.1020)	(0.1012)	(0.1021)	(0.1233)
Ethnic concentration - sq	0.002	0.002	0.002	0.001	0.001	0.014	0.014	0.013	0.011	0.014
	(0.0033)	(0.0033)	(0.0033)	(0.0034)	(0.0038)	(0.0116)	(0.0116)	(0.0115)	(0.0117)	(0.0137)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.007	-0.004
		(0.0038)	(0.0039)	(0.0040)	(0.0042)		(0.0128)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.056	0.056	0.061	0.061	0.145					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year where women turn 15. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at age 15) level.

Table 14: Baseline Results - ITT Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	0.002	0.002	0.002	0.001	0.007	ǂǂ	ǂǂ	-0.001	-0.007	0.009
	(0.0056)	(0.0056)	(0.0056)	(0.0058)	(0.0076)	(0.0185)	(0.0185)	(0.0189)	(0.0194)	(0.0280)
Age at migration		-0.006	-0.005	-0.005	-0.006		-0.009	-0.004	-0.006	-0.007
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0138)	(0.0142)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.054	0.055	0.060	0.061	0.110					
	Panel B: Quadratic specification									
Ethnic concentration	-0.007	-0.007	-0.009	-0.012	-0.007	-0.010	-0.011	-0.021	-0.034	-0.016
	(0.0167)	(0.0167)	(0.0167)	(0.0177)	(0.0192)	(0.0577)	(0.0577)	(0.0580)	(0.0620)	(0.0711)
Ethnic concentration - sq	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.003	0.004	0.004
	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0028)	(0.0082)	(0.0082)	(0.0083)	(0.0089)	(0.0105)
Age at migration		-0.006	-0.006	-0.005	-0.006		-0.009	-0.005	-0.006	-0.007
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0139)	(0.0142)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.055	0.055	0.060	0.061	0.110					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ǂǂis used when rounded estimates are < 0.00

Table 15: Baseline Results - IV Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	0.027	0.027	0.026	0.016	0.175	0.007	0.008	-0.010	-0.069	0.254
	(0.0656)	(0.0655)	(0.0679)	(0.0639)	(0.2263)	(0.2193)	(0.2192)	(0.2303)	(0.2179)	(0.7409)
Age at migration		-0.005	-0.004	-0.005	0.001		-0.009	-0.005	-0.008	0.002
		(0.0043)	(0.0046)	(0.0047)	(0.0095)		(0.0147)	(0.0158)	(0.0162)	(0.0325)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.424	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.047	0.047	0.052	0.057						
F test (1^{st} stage)	30.326	30.437	28.147	29.904	2.823	30.326	30.437	28.147	29.904	2.823
	Panel B: Quadratic specification									
Ethnic concentration	-0.496	-0.511	-0.647	-1.165	-0.062	-0.805	-0.806	-1.674	-3.559	-0.503
	(1.5970)	(1.5914)	(1.5884)	(2.4113)	(2.2502)	(5.0908)	(5.0551)	(4.8472)	(6.5437)	(6.9865)
Ethnic concentration - sq	0.065	0.067	0.084	0.146	0.029	0.101	0.101	0.208	0.433	0.094
	(0.1931)	(0.1924)	(0.1936)	(0.2933)	(0.2538)	(0.6154)	(0.6109)	(0.5904)	(0.7964)	(0.7914)
Age at migration		-0.004	-0.002	-0.001	0.001		-0.007	0.001	0.002	0.005
		(0.0052)	(0.0059)	(0.0076)	(0.0079)		(0.0164)	(0.0179)	(0.0200)	(0.0268)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.054	0.055	0.060	0.061	0.110					
F test (1^{st} stage)	0.699	0.711	0.762	0.439	0.459	0.699	0.711	0.762	0.439	0.459
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports 2SLS estimates for the outcome having taken contraceptives at least once by the age of 20. The right part reports control function estimates from duration analysis where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table 16: Heterogeneity - Immigrants from same origin

	Age at arrival			Length of Stay			Centre Size		
	(3)	(4)	(5)	(3)	(4)	(5)	(3)	(4)	(5)
Panel A: OLS - ITT									
Below mean	ϯϯ	-0.002	0.004	-0.004	-0.004	ϯϯ	0.016	0.006	0.020
	(0.0067)	(0.0069)	(0.0085)	(0.0061)	(0.0063)	(0.0080)	(0.0113)	(0.0136)	(0.0141)
Above mean	0.004	0.004	0.008	0.006	0.005	0.010	0.001	0.001	0.006
	(0.0061)	(0.0063)	(0.0079)	(0.0059)	(0.0061)	(0.0078)	(0.0055)	(0.0057)	(0.0075)
N Obs	4909	4717	4909	4909	4717	4909	4909	4717	4909
p-value equality	0.511	0.419	0.562	0.024	0.031	0.038	0.137	0.682	0.236
Panel B: 2SLS									
Below mean	0.026	0.015	0.173	0.030	0.019	0.176	0.029	0.019	0.172
	(0.0679)	(0.0638)	(0.2251)	(0.0685)	(0.0646)	(0.2284)	(0.0671)	(0.0661)	(0.2146)
Above mean	0.031	0.021	0.182	0.040	0.029	0.185	0.009	0.012	0.145
	(0.0698)	(0.0658)	(0.2322)	(0.0693)	(0.0655)	(0.2275)	(0.0626)	(0.0612)	(0.2018)
N Obs	4909	4717	4909	4909	4717	4909	4909	4717	4909
p-value equality	0.459	0.394	0.362	0.022	0.030	0.083	0.136	0.654	0.200
F test (1^{st} stage)	20.642	22.784	2.207	20.922	22.854	2.257	23.471	24.706	2.539
Panel C: Weibull MLE - ITT									
Below mean	-0.008	-0.014	ϯϯ	-0.021	-0.026	-0.013	0.038	-0.013	0.059
	(0.0226)	(0.0235)	(0.0316)	(0.0212)	(0.0220)	(0.0299)	(0.0388)	(0.0562)	(0.0482)
Above mean	0.003	-0.002	0.013	0.011	0.004	0.020	-0.005	-0.006	0.005
	(0.0208)	(0.0214)	(0.0292)	(0.0198)	(0.0203)	(0.0288)	(0.0188)	(0.0194)	(0.0279)
N Obs	4672	4492	4672	4672	4492	4672	4672	4492	4672
p-value equality	0.602	0.566	0.551	0.035	0.051	0.044	0.236	0.907	0.184
Panel D: Weibull MLE - Control Function									
Below mean	-0.019	-0.077	0.228	-0.008	-0.069	0.250	0.001	-0.075	0.254
	(0.2301)	(0.2169)	(0.7378)	(0.2311)	(0.2185)	(0.7442)	(0.2302)	(0.2277)	(0.7337)
Above mean	-0.007	-0.064	0.250	0.023	-0.039	0.280	-0.049	-0.060	0.176
	(0.2366)	(0.2238)	(0.7590)	(0.2326)	(0.2207)	(0.7410)	(0.2161)	(0.2103)	(0.6917)
N Obs	4672	4492	4672	4672	4492	4672	4672	4492	4672
p-value equality	0.620	0.628	0.507	0.037	0.058	0.059	0.269	0.809	0.218
F test (1^{st} stage)	20.642	22.784	2.207	20.922	22.854	2.257	23.471	24.706	2.539

Note : This table reports various estimates of equation 1.6. Panel A and B reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. Panel C and D reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin, year of arrival dummies and age at migration, columns (3) add household head characteristics at the time of assignment, (4) add the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. In the first three columns concentration is interacted with a dummy being above/below the average age of arrival in the sample. In the following three columns, it is interacted with below/above average length of stay in COA centres and in the last three with below/above average number of asylum seekers registered in the centres at assignment. Panel B reports 2SLS and panel D a control function approach where the interactions of concentration at age 15 are instrumented with the interactions of concentration at assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. F test refers to the refers to the Cragg Donald minimum eigenvalue test. P-value refers to the p-value of the null equality above and below coefficients. ϯϯis used when rounded estimates are < 0.0

Table 17: Results Municipality Level - Naive Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.004	-0.005	-0.005	-0.005	-0.013	-0.017	-0.018	-0.020*	-0.023*	-0.071
	(0.0035)	(0.0035)	(0.0035)	(0.0037)	(0.0147)	(0.0116)	(0.0116)	(0.0116)	(0.0125)	(0.0552)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.006	-0.004
		(0.0038)	(0.0039)	(0.0040)	(0.0042)		(0.0127)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.055	0.056	0.060	0.060	0.145					
	Panel B: Quadratic specification									
Ethnic concentration	-0.009	-0.010	-0.010	-0.009	0.001	-0.033	-0.034	-0.034	-0.031	-0.082
	(0.0108)	(0.0108)	(0.0108)	(0.0109)	(0.0428)	(0.0356)	(0.0356)	(0.0355)	(0.0356)	(0.1623)
Ethnic concentration - sq	††	0.001	††	††	-0.001	0.002	0.002	0.001	0.001	0.001
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0030)	(0.0033)	(0.0033)	(0.0033)	(0.0034)	(0.0107)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.007	-0.004
		(0.0038)	(0.0039)	(0.0039)	(0.0042)		(0.0127)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.055	0.056	0.060	0.060	0.145					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the municipality as of January 1^{st} of the year where women turn 15. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at age 15) level. ††is used when rounded estimates are < 0.00

Table 18: Results Municipality Level - ITT Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.004	-0.004	-0.004	-0.003	-0.008	-0.018	-0.018	-0.017	-0.016	-0.027
	(0.0034)	(0.0034)	(0.0034)	(0.0034)	(0.0116)	(0.0115)	(0.0115)	(0.0115)	(0.0117)	(0.0421)
Age at migration		-0.006	-0.005	-0.005	-0.006		-0.009	-0.004	-0.006	-0.007
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0138)	(0.0142)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.055	0.055	0.060	0.061	0.110					
	Panel B: Quadratic specification									
Ethnic concentration	0.004	0.004	0.004	ŧŧ	0.002	0.006	0.006	0.007	-0.006	0.039
	(0.0098)	(0.0098)	(0.0098)	(0.0099)	(0.0262)	(0.0333)	(0.0334)	(0.0333)	(0.0335)	(0.0927)
Ethnic concentration - sq	-0.001	-0.001	-0.001	ŧŧ	-0.001	-0.003	-0.003	-0.003	-0.001	-0.006
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0022)	(0.0035)	(0.0035)	(0.0035)	(0.0036)	(0.0080)
Age at migration		-0.006	-0.005	-0.005	-0.006		-0.009	-0.004	-0.006	-0.007
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0139)	(0.0142)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.424	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.055	0.055	0.060	0.061	0.110					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the municipality as of January 1^{st} of the year of assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ŧŧis used when rounded estimates are < 0.00

Table 19: Results Municipality Level - IV Estimates - Immigrants from the same origin

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.021	-0.020	-0.021	-0.016	-0.033	-0.089	-0.088	-0.089	-0.080	-0.109
	(0.0174)	(0.0173)	(0.0175)	(0.0174)	(0.0472)	(0.0577)	(0.0572)	(0.0584)	(0.0590)	(0.1754)
Age at migration		-0.007*	-0.007*	-0.007	-0.009*		-0.015	-0.012	-0.013	-0.016
		(0.0040)	(0.0042)	(0.0043)	(0.0053)		(0.0136)	(0.0144)	(0.0148)	(0.0190)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.424	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.049	0.051	0.055	0.059	0.099					
F test (1^{st} stage)	85.710	87.443	83.886	82.805	18.082	85.710	87.443	83.886	82.805	18.082
	Panel B: Quadratic specification									
Ethnic concentration	0.025	0.028	0.032	-0.001	0.073	0.042	0.042	0.055	-0.060	0.573
	(0.0763)	(0.0757)	(0.0757)	(0.0742)	(0.2427)	(0.2593)	(0.2589)	(0.2609)	(0.2494)	(0.9070)
Ethnic concentration - sq	-0.004	-0.005	-0.005	-0.001	-0.009	-0.013	-0.013	-0.014	-0.002	-0.059
	(0.0066)	(0.0066)	(0.0067)	(0.0065)	(0.0206)	(0.0232)	(0.0232)	(0.0236)	(0.0226)	(0.0785)
Age at migration		-0.007*	-0.007	-0.006	-0.007		-0.014	-0.011	-0.013	-0.008
		(0.0041)	(0.0043)	(0.0044)	(0.0061)		(0.0141)	(0.0149)	(0.0152)	(0.0217)
N Obs	4909	4909	4909	4717	4909	4672	4672	4672	4492	4672
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.395	0.397
R squared	0.054	0.055	0.060	0.061	0.110					
F test (1^{st} stage)	53.201	53.730	52.747	54.296	3.585	53.201	53.730	52.747	54.296	3.585
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports 2SLS estimates for the outcome having taken contraceptives at least once by the age of 20. The right part reports control function estimates from duration analysis where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of immigrants from the same origin (first and second generations) living in the municipality as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table 20: Increasing sample size - Naive Estimates

	Peers from the same origin					All Non Western Immigrants				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
					Panel A: Linear specification					
Ethnic concentration	-0.013**	-0.015**	-0.016**	-0.014**	-0.015**	-0.007	-0.011*	-0.012**	-0.009	-0.022**
	(0.0056)	(0.0056)	(0.0055)	(0.0061)	(0.0077)	(0.0058)	(0.0058)	(0.0058)	(0.0066)	(0.0098)
Age at migration		-0.016***	-0.015***	-0.015***	-0.015***		-0.016***	-0.015***	-0.015***	-0.016***
		(0.0025)	(0.0027)	(0.0027)	(0.0028)		(0.0025)	(0.0027)	(0.0027)	(0.0028)
N Obs	6794	6794	6794	6771	6794	6794	6794	6794	6771	6794
Mean Outcome	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400
R squared	0.051	0.057	0.062	0.062	0.120	0.050	0.056	0.061	0.061	0.120
					Panel B: Quadratic specification					
Ethnic concentration	-0.039*	-0.039	-0.035	-0.033	-0.027	-0.044	-0.041	-0.033	-0.039	0.011
	(0.0237)	(0.0236)	(0.0234)	(0.0238)	(0.0265)	(0.0409)	(0.0411)	(0.0408)	(0.0445)	(0.0527)
Ethnic concentration - sq	0.003	0.003	0.002	0.002	0.001	0.003	0.002	0.002	0.002	-0.003
	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0030)	(0.0031)	(0.0031)	(0.0031)	(0.0034)	(0.0042)
Age at migration		-0.016***	-0.015***	-0.015***	-0.015***		-0.016***	-0.015***	-0.015***	-0.016***
		(0.0025)	(0.0027)	(0.0027)	(0.0028)		(0.0025)	(0.0027)	(0.0027)	(0.0028)
N Obs	6794	6794	6794	6771	6794	6794000	6794000	6794000	6771000	6794000
Mean Outcome	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400
R squared	0.051	0.057	0.062	0.062	0.120	0.050	0.056	0.061	0.061	0.120
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turns 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. On the left part of the table ethnic concentration is measured as the log of the number of immigrants from the same region and on the right part of the table as the log of non-Western immigrants (in both cases, first and second generations) living in the zip 4 as of January 1^{st} of the year where women turn 15. The sample only includes complete and incomplete spells. Standard errors are clustered at the year of arrival/municipality (at age 15) level.

Table 21: Increasing sample size - ITT Estimates

	Peers from the same origin					All Non Western Immigrants				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	0.005	0.005	0.004	0.004	0.006	0.001	0.001	++	0.001	0.005
	(0.0045)	(0.0045)	(0.0045)	(0.0046)	(0.0063)	(0.0048)	(0.0047)	(0.0047)	(0.0049)	(0.0095)
Age at migration		-0.015***	-0.014***	-0.014***	-0.014***		-0.016***	-0.014***	-0.014***	-0.014***
		(0.0026)	(0.0028)	(0.0028)	(0.0028)		(0.0026)	(0.0028)	(0.0028)	(0.0028)
N Obs	6866	6866	6866	6588	6866	6851	6851	6851	6588	6851
Mean Outcome	0.400	0.400	0.400	0.398	0.400	0.400	0.400	0.400	0.398	0.400
R squared	0.050	0.056	0.061	0.062	0.097	0.050	0.056	0.060	0.061	0.097
	Panel B: Quadratic specification									
Ethnic concentration	-0.007	-0.007	-0.009	-0.012	-0.007	0.019	0.017	0.010	0.014	0.030
	(0.0134)	(0.0135)	(0.0135)	(0.0148)	(0.0158)	(0.0235)	(0.0234)	(0.0232)	(0.0309)	(0.0333)
Ethnic concentration - sq	0.002	0.002	0.002	0.003	0.002	-0.002	-0.001	-0.001	-0.001	-0.003
	(0.0020)	(0.0020)	(0.0020)	(0.0022)	(0.0024)	(0.0021)	(0.0021)	(0.0021)	(0.0027)	(0.0033)
Age at migration		-0.015***	-0.014***	-0.014***	-0.014***		-0.016***	-0.014***	-0.014***	-0.014***
		(0.0026)	(0.0028)	(0.0028)	(0.0028)		(0.0026)	(0.0028)	(0.0028)	(0.0028)
N Obs	6866	6866	6866	6588	6866	6851	6851	6851	6588	6851
Mean Outcome	0.400	0.400	0.400	0.398	0.400	0.400	0.400	0.400	0.398	0.400
R squared	0.050	0.056	0.061	0.062	0.097	0.050	0.056	0.060	0.061	0.097
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. On the left part of the table ethnic concentration is measured as the log of the number of immigrants from the same region and on the right part of the table as the log of non-Western immigrants (in both cases, first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. The sample only includes complete and incomplete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ++is used when rounded estimates are < 0.00

Table 22: Increasing sample size - IV Estimates

	Peers from the same origin					All Non Western Immigrants				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
					Panel A: Linear specification					
Ethnic concentration	0.056	0.053	0.048	0.043	0.175	0.008	0.004	0.004	0.010	-0.185
	(0.0513)	(0.0515)	(0.0529)	(0.0514)	(0.2024)	(0.0380)	(0.0383)	(0.0394)	(0.0367)	(0.3663)
Age at migration		-0.014***	-0.012***	-0.013***	-0.008		-0.015***	-0.014***	-0.014***	-0.024
		(0.0032)	(0.0035)	(0.0035)	(0.0082)		(0.0033)	(0.0035)	(0.0034)	(0.0192)
N Obs	6866	6866	6866	6588	6866	6851	6851	6851	6588	6851
Mean Outcome	0.400	0.400	0.400	0.398	0.400	0.400	0.400	0.400	0.398	0.400
R squared	0.025	0.031	0.040	0.044		0.049	0.055	0.060	0.060	
F test (1^{st} stage)	42.997	42.621	39.682	40.163	3.529	66.778	65.141	61.388	74.004	1.165
					Panel B: Quadratic specification					
Ethnic concentration	-0.364	-0.371	-0.512	-1114	0.603	0.459	0.404	0.221	0.324	0.551
	-12713	-12725	-11767	-26540	-66794	(0.6445)	(0.6288)	(0.5498)	(0.6725)	-11154
Ethnic concentration - sq	0.052	0.052	0.070	0.144	-0.044	-0.036	-0.032	-0.018	-0.026	-0.061
	(0.1529)	(0.1532)	(0.1422)	(0.3249)	(0.7090)	(0.0501)	(0.0489)	(0.0427)	(0.0537)	(0.0908)
Age at migration		-0.012**	-0.010*	-0.007	-0.006		-0.016***	-0.014***	-0.014***	-0.026
		(0.0042)	(0.0051)	(0.0118)	(0.0162)		(0.0033)	(0.0034)	(0.0035)	(0.0207)
N Obs	6866	6866	6866	6588	6866	6851000	6851000	6851000	6588000	6851000
Mean Outcome	0.400	0.400	0.400	0.398	0.400	0.400	0.400	0.400	0.398	0.400
R squared								0.043	0.023	
F test (1^{st} stage)	0.802	0.804	1.014	0.305	0.038	2.915	2.895	3.413	2.861	0.727
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports 2SLS estimates for the outcome having taken contraceptives at least once by the age of 20. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. On the left part of the table ethnic concentration is measured as the log of the number of immigrants from the same region and on the right part of the table as the log of non-Western immigrants. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table 23: Falsification Test - Educational Outcomes - Naive Estimates - Immigrants from the same origin

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	0.003	0.004	0.006	0.008	0.006	0.003	0.002	0.004	0.004	0.007
	(0.0069)	(0.0069)	(0.0066)	(0.0069)	(0.0093)	(0.0055)	(0.0055)	(0.0053)	(0.0057)	(0.0073)
Age at migration		0.010**	0.010**	0.009**	0.012**		-0.005**	-0.006**	-0.006**	-0.005**
		(0.0040)	(0.0040)	(0.0041)	(0.0042)		(0.0022)	(0.0023)	(0.0023)	(0.0024)
N Obs	4592	4592	4592	4581	4592	6955	6955	6955	6921	6955
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.158	0.244	0.091	0.092	0.136	0.135	0.192
	Panel B: Quadratic specification									
Ethnic concentration	0.020	0.021	0.023	0.013	0.016	ŧŧ	ŧŧ	-0.001	-0.008	-0.015
	(0.0302)	(0.0301)	(0.0286)	(0.0294)	(0.0321)	(0.0242)	(0.0243)	(0.0229)	(0.0234)	(0.0260)
Ethnic concentration- sq	-0.002	-0.002	-0.002	-0.001	-0.001	ŧŧ	ŧŧ	0.001	0.001	0.003
	(0.0034)	(0.0034)	(0.0032)	(0.0033)	(0.0037)	(0.0027)	(0.0028)	(0.0026)	(0.0027)	(0.0030)
Age at migration		0.010**	0.010**	0.009**	0.012**		-0.005**	-0.006**	-0.006**	-0.005**
		(0.0040)	(0.0040)	(0.0041)	(0.0042)		(0.0022)	(0.0023)	(0.0023)	(0.0024)
N Obs	4592	4592	4592	4581	4592	6955	6955	6955	6921	6955
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.158	0.244	0.091	0.092	0.136	0.135	0.192
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year where women turn 15. Standard errors are clustered at the year of arrival/municipality (at age 15) level. ŧŧis used when rounded estimates are < 0.00

Table 24: Falsification Test - Educational Outcomes - ITT Estimates - Immigrants from the same origin

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.005	-0.005	-0.006	-0.007	-0.003	-0.006	-0.006	-0.006	-0.008*	-0.007
	(0.0055)	(0.0055)	(0.0054)	(0.0057)	(0.0076)	(0.0045)	(0.0045)	(0.0044)	(0.0044)	(0.0061)
Age at migration		0.010**	0.010**	0.009**	0.011**		-0.005**	-0.006**	-0.006**	-0.006**
		(0.0038)	(0.0038)	(0.0039)	(0.0039)		(0.0022)	(0.0023)	(0.0024)	(0.0024)
N Obs	4601	4601	4601	4421	4601	7038	7038	7038	6754	7038
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.162	0.202	0.092	0.093	0.136	0.138	0.165
	Panel B: Quadratic specification									
Ethnic concentration	-0.010	-0.009	-0.014	-0.015	-0.009	-0.014	-0.014	-0.016	-0.021	-0.008
	(0.0159)	(0.0158)	(0.0157)	(0.0163)	(0.0186)	(0.0131)	(0.0131)	(0.0128)	(0.0134)	(0.0148)
Ethnic concentration- sq	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.000
	(0.0024)	(0.0024)	(0.0023)	(0.0024)	(0.0027)	(0.0020)	(0.0020)	(0.0019)	(0.0020)	(0.0022)
Age at migration		0.010**	0.010**	0.009**	0.011**		-0.005**	-0.006**	-0.006**	-0.006**
		(0.0038)	(0.0038)	(0.0039)	(0.0039)		(0.0022)	(0.0023)	(0.0024)	(0.0024)
N Obs	4601	4601	4601	4421	4601	7038	7038	7038	6754	7038
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.162	0.202	0.092	0.093	0.136	0.138	0.165
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the time of assignment. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ††is used when rounded estimates are < 0.00

Table 25: Falsification Test - Educational Outcomes - IV Estimates - Immigrants from the same origin

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.064	-0.065	-0.078	-0.078	-0.086	-0.057	-0.059	-0.063	-0.074*	-0.138
	(0.0669)	(0.0668)	(0.0680)	(0.0657)	(0.1963)	(0.0432)	(0.0435)	(0.0437)	(0.0427)	(0.1367)
Age at migration		0.008*	0.007	0.006	0.008		-0.007**	-0.008**	-0.008**	-0.011**
		(0.0045)	(0.0048)	(0.0049)	(0.0089)		(0.0026)	(0.0028)	(0.0029)	(0.0054)
N Obs	4601	4601	4601	4421	4601	7038	7038	7038	6754	7038
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.094	0.095	0.122	0.126	0.160	0.069	0.070	0.109	0.101	0.050
F test (1^{st} stage)	26.289	26.407	24.824	26.394	2.651	56.513	55.652	52.603	53.412	6.247
	Panel B: Quadratic specification									
Ethnic concentration	-0.922	-0.879	-1.167	-1.552	-0.952	-1.009	-1.008	-1.088	-1.744	-398.487
	(1.3031)	(1.2811)	(1.3126)	(1.8635)	(1.9398)	(0.9903)	(0.9910)	(0.9347)	(1.5724)	(121090.3443)
Ethnic concentration- sq	0.106	0.101	0.136	0.183	0.104	0.116	0.116	0.125	0.205	42.580
	(0.1579)	(0.1551)	(0.1600)	(0.2275)	(0.2177)	(0.1174)	(0.1176)	(0.1108)	(0.1893)	(12940.5574)
Age at migration		0.009*	0.010*	0.010	0.009		-0.003	-0.004	0.000	-0.426
		(0.0051)	(0.0059)	(0.0071)	(0.0086)		(0.0043)	(0.0046)	(0.0082)	-1270011
N Obs	4601	4601	4601	4421	4601	7038	7038	7038	6754	7038
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared										
F test (1^{st} stage)	0.870	0.881	0.982	0.628	0.467	1.700	1.697	1.991	1.106	††
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ††is used when rounded estimates are < 0.00

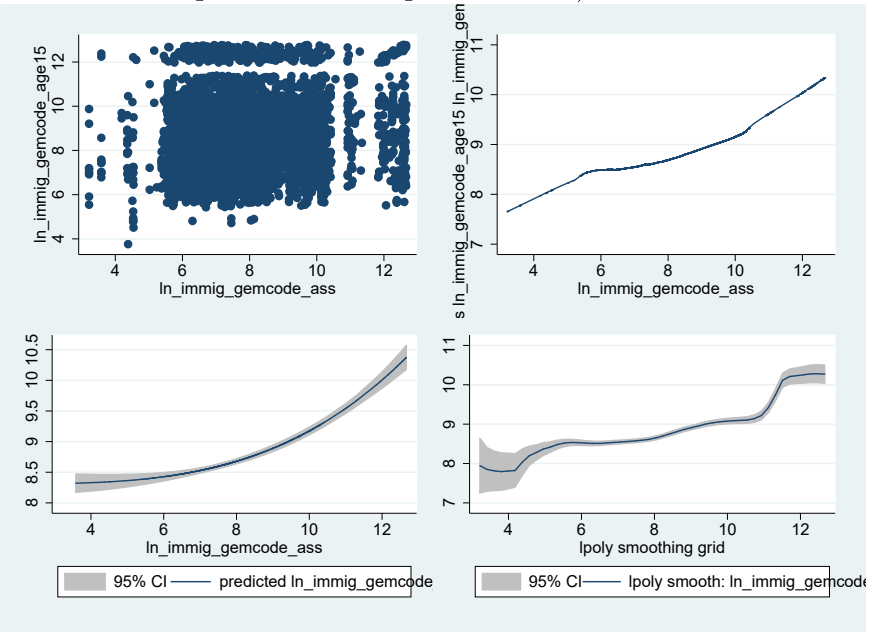
Table 26: Falsification test - Heterogeneity - Immigrants from same origin

	Age at arrival			Length of Stay			Centre Size		
	(3)	(4)	(5)	(3)	(4)	(5)	(3)	(4)	(5)
Panel A: OLS - ITT - Complete Spells									
Below mean	-0.004	-0.004	-0.001	-0.006	-0.006	-0.005	-0.013*	-0.015*	-0.009
	(0.0061)	(0.0064)	(0.0083)	(0.0063)	(0.0066)	(0.0083)	(0.0077)	(0.0091)	(0.0120)
Above mean	-0.008	-0.009	-0.005	-0.006	-0.007	-0.003	-0.006	-0.006	-0.003
	(0.0061)	(0.0063)	(0.0081)	(0.0056)	(0.0059)	(0.0079)	(0.0055)	(0.0057)	(0.0077)
p-value	0.571	0.432	0.494	0.995	0.855	0.689	0.300	0.280	0.511
N obs	4601	4421	4601	4601	4421	4601	4601	4421	4601
Panel B: 2SLS - Complete Spells									
Below mean	-0.077	-0.077	-0.084	-0.078	-0.079	-0.086	-0.080	-0.085	-0.085
	(0.0679)	(0.0655)	(0.1949)	(0.0680)	(0.0658)	(0.1961)	(0.0677)	(0.0666)	(0.1929)
Above mean	-0.082	-0.084	-0.091	-0.078	-0.080	-0.084	-0.065	-0.068	-0.071
	(0.0698)	(0.0675)	(0.1996)	(0.0682)	(0.0664)	(0.1951)	(0.0637)	(0.0629)	(0.1780)
p-value	0.439	0.320	0.414	0.939	0.739	0.612	0.113	0.118	0.445
N obs	4601	4421	4601	4601	4421	4601	4601	4421	4601
F test	19.311	21.138	2.188	19.402	21.029	2.233	22.175	23.116	2.631
Panel C: OLS - ITT - Complete and incomplete Spells									
Below mean	-0.005	-0.006	-0.005	-0.006	-0.007	-0.006	-0.014**	-0.020**	-0.010
	(0.0053)	(0.0053)	(0.0069)	(0.0053)	(0.0054)	(0.0068)	(0.0070)	(0.0081)	(0.0099)
Above mean	-0.008	-0.009*	-0.008	-0.007	-0.009*	-0.007	-0.006	-0.007	-0.006
	(0.0048)	(0.0049)	(0.0064)	(0.0045)	(0.0045)	(0.0063)	(0.0044)	(0.0045)	(0.0061)
p-value	0.523	0.517	0.542	0.729	0.735	0.810	0.177	0.106	0.678
N obs	7038	6754	7038	7038	6754	7038	7038	6754	7038
Panel D: 2SLS - Complete and incomplete Spells									
Below mean	-0.061	-0.072*	-0.137	-0.064	-0.075*	-0.139	-0.070	-0.087**	-0.141
	(0.0433)	(0.0423)	(0.1368)	(0.0437)	(0.0428)	(0.1373)	(0.0446)	(0.0440)	(0.1372)
Above mean	-0.067	-0.078*	-0.145	-0.066	-0.078*	-0.141	-0.055	-0.067	-0.128
	(0.0445)	(0.0436)	(0.1412)	(0.0440)	(0.0433)	(0.1381)	(0.0421)	(0.0415)	(0.1282)
p-value	0.323	0.294	0.274	0.585	0.520	0.713	0.072	0.035	0.311
N obs	7038	6754	7038	7038	6754	7038	7038	6754	7038
F test	43.796	45.109	4.839	42.918	43.866	4.905	46.995	47.420	5.503

Note : This table reports various estimates of equation 1.6. The outcome variable is having at least attended HAVO as highest education level of education. All specifications control for country of origin, year of arrival dummies and age at migration, columns (3) add household head characteristics at the time of assignment, (4) add the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. In the first three columns concentration is interacted with a dummy being above/below the average age of arrival in the sample. In the following three columns, it s interacted with below/above average length of stay in COA centres and in the last three with below/above average number of asylum seekers registered in the centres at assignment. Panel B and D report 2SLS where the interactions of concentration at age 15 are instrumented with the interactions of concentration at assignment. Panel A and B include complete spells, panel C and D complete and incomplete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. F test refers to the refers to the Cragg Donald minimum eigenvalue test. P-value refers to the p-value of the null equality above and below coefficients.

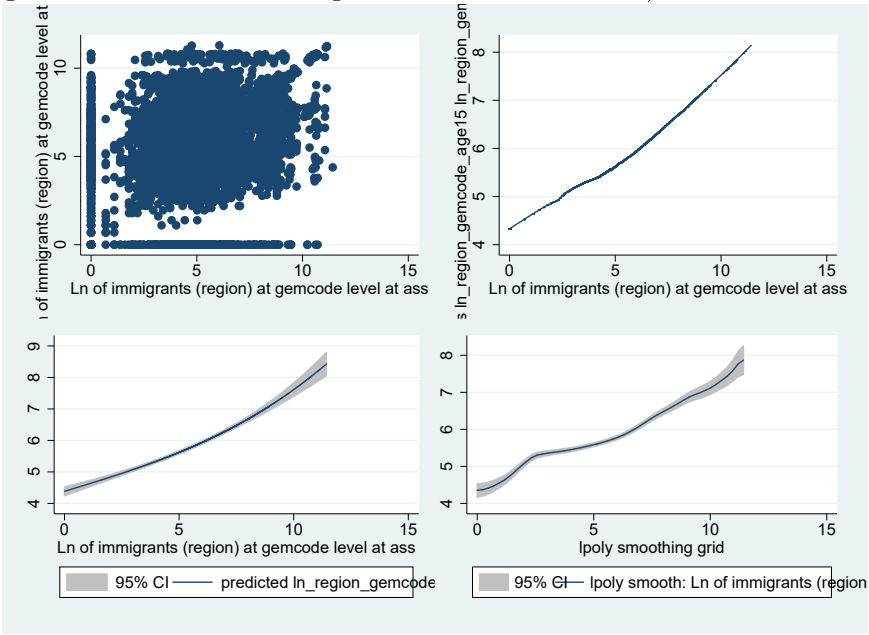
10 Appendix

Figure A1: Community size at assignment and age 15 (non-Western immigrants, zip4 neighbourhoods)



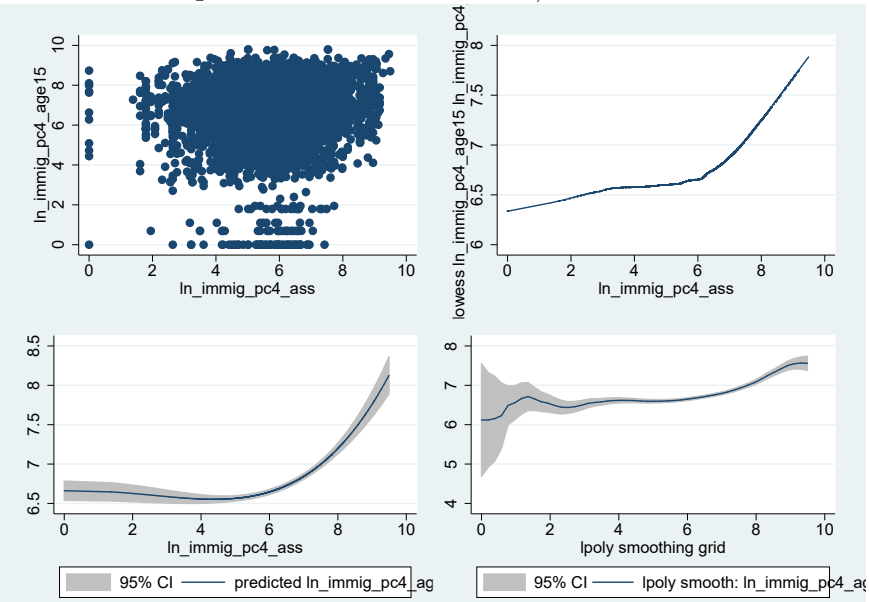
Note : I plot the log of ethnic concentration at age 15 against the log of ethnic concentration at the time of assignment. Ethnic concentration is defined as immigrants from non Western countries and neighbourhood refers to municipalities. The figure is split into four quadrants, in the upper left one, I scatter the points, in the upper right one I approximate the scatter of points with a locally weighted regressions, in the lower left one I use a quadratic approximation and in the lower right one a kernel-weighted local polynomial.

Figure A2: Community size at assignment and age 15 (Immigrants from the same region, municipality level)



Note : I plot the log of ethnic concentration at age 15 against the log of ethnic concentration at the time of assignment. Ethnic concentration is defined as immigrants from the same region and neighbourhood refers to municipalities. The figure is split into four quadrants, in the upper left one, I scatter the points, in the upper right one I approximate the scatter of points with a locally weighted regressions, in the lower left one I use a quadratic approximation and in the lower right one a kernel-weighted local polynomial.

Figure A3: Community size at assignment and age 15 (non-Western immigrants, municipality level)



Note : I plot the log of ethnic concentration at age 15 against the log of ethnic concentration at the time of assignment. Ethnic concentration is defined as immigrants from non Western countries and neighbourhood refers to zip4. The figure is split into four quadrants, in the upper left one, I scatter the points, in the upper right one I approximate the scatter of points with a locally weighted regressions, in the lower left one I use a quadratic approximation and in the lower right one a kernel-weighted local polynomial.

Table A1: Differences in contraceptive use by year

	Age 15	Age 16	Age 17	Age 18	Age 19	Age 20
2007	0.00470					
	(0.00673)					
2008	0.00565	0.000814				
	(0.00679)	(0.00958)				
2009	0.0106	0.00498	0.00562			
	(0.00674)	(0.00967)	(0.0114)			
2010	0.0180**	0.0149	0.0206*	0.0102		
	(0.00683)	(0.00959)	(0.0115)	(0.0123)		
2011	0.0226**	0.0116	0.0266**	0.0218*	-0.0000116	
	(0.00688)	(0.00972)	(0.0114)	(0.0124)	(0.0127)	
2012	0.0239***	0.0248**	0.0174	0.0284**	0.0226*	-0.00247
	(0.00703)	(0.00980)	(0.0116)	(0.0123)	(0.0129)	(0.0128)
2013		0.0223**	0.0323**	0.0159	0.0263**	0.0217*
		(0.0100)	(0.0117)	(0.0125)	(0.0127)	(0.0130)
2014			0.0271**	0.0416***	0.0155	0.0142
			(0.0119)	(0.0126)	(0.0129)	(0.0129)
2015				0.0293**	0.0298**	0.00923
				(0.0129)	(0.0130)	(0.0130)
2016					0.0139	0.0235*
					(0.0133)	(0.0131)
2017						-0.00440
						(0.0134)
Constant	0.0546***	0.141***	0.237***	0.332***	0.420***	0.496***
	(0.00486)	(0.00692)	(0.00823)	(0.00889)	(0.00920)	(0.00928)
Country FE	✓	✓	✓	✓	✓	✓
N Obs	19344	19215	19187	19152	19168	19344
F test	3.653	2.056	2.074	2.382	1.747	1.597
p-value	0.00126	0.0550	0.0529	0.0266	0.106	0.143

Note : Each column represents the coefficients of a linear regression where the outcome variable is a dummy for having taken contraceptives at least once by a certain age. The explanatory variables are country of origin fixed effects and year of usage dummies. In the last two rows, I report (i) the value of the F test where the null hypothesis is the equality of the coefficients on year of usage dummies and (ii) the p-value of the test.

Table A2: Grouping countries into regions

Country	Nb obs	% of the sample
Afghanistan	2,127	24.56
Iraq	1,509	17.42
Soviet Union	631	7.29
Iran	567	6.55
Yugoslavia	551	6.36
Somalia	528	6.10
Syria	303	3.50
Angola	232	2.68
Azerbaijan	206	2.38
Russia	200	2.31
Turkey	174	2.01
FR Yugoslavia	156	1.80
Armenia	148	1.71
Bosnia Herzegovina	128	1.48
Sudan	126	1.45
Comores	122	1.41
Ethiopia	67	0.77
Burundi	67	0.77
Pakistan	50	0.58
Sri Lanka	49	0.57
Rwanda	47	0.54
Ukraine	44	0.51
Kuwait	41	0.47
Georgia	36	0.42
Colombia	33	0.38
Sierra Leone	25	0.29
Lebanon	25	0.29
Algeria	24	0.28
Saudi Arabia	23	0.27

Note : This table shows the distribution of origin countries for asylum seekers who are part of the experimental population (complete and incomplete spells).

Table A3: Variation in Community sizes (Afghanistan)

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	4	14	10	37	54	189	512	813
25th p	13	31	40	78	134	482	829	1694
Median	30	59	81	226	326	1027	1611	5908
75th p	62	119	173	838	757	2038	5595	20264
90th p	101	214	626	2249	1481	3342	20735	44134
Mean	46	93	494	1091	679	1562	15028	28180
Nb Obs	2127	2127	2127	2127	2123	2127	1957	2004

Note : This table reports, for asylum seekers from Afghanistan, the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table A4: Variation in Community sizes (Iraq)

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	5	19	15	40	65	209	642	890
25th p	21	51	76	160	186	491	1228	1886
Median	61	137	342	938	458	1028	2516	6558
75th p	158	373	1471	3977	972	2046	11024	23000
90th p	387	916	6143	11305	1668	3287	24524	54175
Mean	158	322	2691	4932	795	1513	15504	28178
Nb Obs	1509	1509	1509	1509	1506	1509	1401	1410

Note : This table reports, for asylum seekers from Iraq, the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table A5: Variation in Community sizes (Iran)

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	4	12	8	27	65	206	613	916
25th p	14	29	41	80	179	480	1049	1818
Median	30	54	95	218	401	932	2307	6013
75th p	63	105	223	868	779	1855	7733	22260
90th p	101	172	671	2139	1474	3085	21173	54162
Mean	45	80	529	1037	695	1396	16133	27081
Nb Obs	567	567	567	567	566	567	521	522

Note : This table reports, for asylum seekers from Iran, the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table A6: Variation in Community sizes (Yougoslavia)

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	8	15	33	22	70	115	561	920
25th p	17	42	84	126	155	394	961	1800
Median	45	90	190	412	405	830	1845	5285
75th p	92	135	705	1210	847	1526	7395	14244
90th p	163	217	1550	2030	1297	2912	19215	31688
Mean	73	108	892	1313	693	1316	13353	21075
Nb Obs	551	551	551	551	548	551	508	499

Note : This table reports, for asylum seekers from Yugoslavia, the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table A7: Variation in Community sizes (Soviet Union)

	Same region				non-Western immigrants			
	At assignment		At age 15		At assignment		At age 15	
	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality	Zip4	Municipality
10th p	5	20	14	47	68	160	564	837
25th p	14	42	67	113	164	365	942	1382
Median	36	73	119	219	404	663	1779	3033
75th p	62	106	271	633	750	1307	5595	14079
90th p	97	144	773	1455	1233	2863	18930	28769
Mean	44	80	338	677	647	1204	10804	21363
Nb Obs	631	631	631	631	630	631	579	579

Note : This table reports, for asylum seekers from the Soviet Union, the number of immigrants (from first or second generation, staying at a COA centre or not) who lived in the same zip4 or municipality as of January 1^{st} of the year of assignment and the year a young woman turned 15. I report community sizes of immigrants from the same region and all non-Western immigrants in the zip4 and municipality.

Table A8: Baseline Results - Naive Estimates - Immigrants from non Western countries

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.007	-0.008	-0.009	-0.008	-0.013	-0.036	-0.038	-0.042*	-0.044*	-0.059
	(0.0069)	(0.0069)	(0.0069)	(0.0079)	(0.0117)	(0.0235)	(0.0236)	(0.0237)	(0.0267)	(0.0414)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.007	-0.004
		(0.0038)	(0.0039)	(0.0040)	(0.0042)		(0.0128)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.055	0.056	0.060	0.060	0.145					
	Panel B: Quadratic specification									
Ethnic concentration	-0.029	-0.028	-0.020	-0.012	0.034	-0.174*	-0.176*	-0.174*	-0.155	-0.186
	(0.0472)	(0.0473)	(0.0472)	(0.0527)	(0.0601)	(0.1018)	(0.1020)	(0.1012)	(0.1021)	(0.1233)
Ethnic concentration - sq	0.002	0.002	0.001	0.000	-0.004	0.014	0.014	0.013	0.011	0.014
	(0.0036)	(0.0036)	(0.0036)	(0.0041)	(0.0048)	(0.0116)	(0.0116)	(0.0115)	(0.0117)	(0.0137)
Age at migration		-0.007*	-0.006	-0.006	-0.005		-0.011	-0.007	-0.007	-0.004
		(0.0038)	(0.0039)	(0.0040)	(0.0042)		(0.0128)	(0.0134)	(0.0135)	(0.0152)
N Obs	4897	4897	4897	4886	4897	4660	4660	4660	4650	4660
Mean Outcome	0.427	0.427	0.427	0.426	0.427	0.397	0.397	0.397	0.397	0.397
R squared	0.055	0.056	0.060	0.060	0.145					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year where women turn 15. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at age 15) level.

Table A9: Baseline Results - ITT Estimates - Immigrants from non Western countries

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.001	-0.002	-0.001	-0.002	0.014	-0.013	-0.013	-0.012	-0.018	0.022
	(0.0060)	(0.0060)	(0.0060)	(0.0064)	(0.0118)	(0.0201)	(0.0201)	(0.0204)	(0.0213)	(0.0435)
Age at migration		-0.006*	-0.006	-0.005	-0.006		-0.010	-0.006	-0.006	-0.008
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0138)	(0.0142)
N Obs	4901	4901	4901	4717	4901	4664	4664	4664	4492	4646
Mean Outcome	0.426	0.426	0.426	0.424	0.426	0.397	0.397	0.397	0.397	0.397
R squared	0.054	0.055	0.059	0.061	0.110					
	Panel B: Quadratic specification									
Ethnic concentration	0.036	0.036	0.028	0.029	0.019	0.156	0.156	0.126	0.101	0.118
	(0.0290)	(0.0290)	(0.0291)	(0.0378)	(0.0408)	(0.1123)	(0.1125)	(0.1131)	(0.1339)	(0.1740)
Ethnic concentration - sq	-0.003	-0.003	-0.003	-0.003	-0.001	-0.015	-0.015	-0.012	-0.011	-0.010
	(0.0026)	(0.0026)	(0.0026)	(0.0033)	(0.0039)	(0.0097)	(0.0097)	(0.0098)	(0.0116)	(0.0164)
Age at migration		-0.006*	-0.006	-0.005	-0.006		-0.010	-0.006	-0.006	-0.008
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0128)	(0.0135)	(0.0138)	(0.0142)
N Obs	4901	4901	4901	4717	4901	4664	4664	4664	4492	4664
Mean Outcome	0.426	0.426	0.426	0.424	0.426	0.397	0.397	0.397	0.395	0.397
R squared	0.055	0.055	0.060	0.061	0.110					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table A10: Baseline Results - IV Estimates - Immigrants from non Western countries

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.013	-0.013	-0.011	-0.013	-1.994	-0.121	-0.121	-0.118	-0.154	-3.220
	(0.0525)	(0.0526)	(0.0552)	(0.0530)	(8.6218)	(0.1784)	(0.1787)	(0.1896)	(0.1787)	(6.4044)
Age at migration		-0.007	-0.006	-0.006	-0.103		-0.016	-0.012	-0.014	-0.164
		(0.0044)	(0.0047)	(0.0048)	(0.4179)		(0.0150)	(0.0160)	(0.0164)	(0.3080)
N Obs	4901	4901	4901	4717	4901	4664	4664	4664	4492	4664
Mean Outcome	0.426	0.426	0.426	0.424	0.426	0.397	0.397	0.397	0.395	0.397
R squared	0.054	0.055	0.060	0.061						
F test (1^{st} stage)	43.405	42.774	39.079	47.197	0.050	43.405	42.774	39.079	47.197	0.050
	Panel B: Quadratic specification									
Ethnic concentration	1.057	1.035	0.761	0.606	2.062	4.009	3.976	2.820	1.219	1.973
	(1.2023)	(1.1996)	(1.0829)	(1.1364)	(2.6932)	(4.1713)	(4.1865)	(4.0193)	(3.8463)	(7.3018)
Ethnic concentration - sq	-0.083	-0.081	-0.060	-0.048	-0.175	-0.319	-0.316	-0.228	-0.107	-0.228
	(0.0906)	(0.0904)	(0.0816)	(0.0864)	(0.2100)	(0.3145)	(0.3156)	(0.3030)	(0.2928)	(0.5816)
Age at migration		-0.007	-0.007	-0.006	-0.017		-0.016	-0.013	-0.015	-0.056
		(0.0048)	(0.0048)	(0.0048)	(0.0231)		(0.0150)	(0.0157)	(0.0158)	(0.0663)
N Obs	4901	4901	4901	4717	4901	4672	4672	4672	4492	4672
Mean Outcome	0.426	0.426	0.426	0.424	0.426	0.397	0.397	0.397	0.395	0.397
R squared				0.005						
F test (1^{st} stage)	2.290	2.277	2.490	2.589	0.775	2.290	2.277	2.490	2.589	0.775
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports 2SLS estimates for the outcome having taken contraceptives at least once by the age of 20. The right part reports control function estimates from duration analysis where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of non-Western immigrants(first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table A11: Heterogeneity - Immigrants from non Western countries

	Age at arrival			Length of Stay			Centre Size		
	(3)	(4)	(5)	(3)	(4)	(5)	(3)	(4)	(5)
Panel A: OLS - ITT									
Below mean	-0.003	-0.004	0.012	-0.004	-0.004	0.011	0.009	0.001	0.021
	(0.0065)	(0.0069)	(0.0122)	(0.0061)	(0.0066)	(0.0119)	(0.0089)	(0.0102)	(0.0138)
Above mean	0.000	-0.000	0.014	0.001	0.001	0.016	-0.003	-0.002	0.012
	(0.0063)	(0.0066)	(0.0118)	(0.0062)	(0.0066)	(0.0119)	(0.0060)	(0.0064)	(0.0117)
N Obs	4901	4717	4901	4901	4717	4901	4901	4717	4901
p-value equality	0.475	0.430	0.548	0.072	0.101	0.067	0.119	0.693	0.245
Panel B: 2SLS									
Below mean	-0.011	-0.013	-1.995	-0.005	-0.008	-2.980	-0.011	-0.012	-6.716
	(0.0552)	(0.0528)	(8.6579)	(0.0545)	(0.0523)	(18.2904)	(0.0551)	(0.0539)	(110.9624)
Above mean	-0.009	-0.010	-2.010	-0.001	-0.003	-2.926	-0.022	-0.015	-6.536
	(0.0559)	(0.0537)	(8.7312)	(0.0541)	(0.0520)	(17.9913)	(0.0531)	(0.0514)	(107.8002)
N Obs	4901	4717	4901	4786	4601	4575	4901	4717	4901
p-value equality	0.495	0.455	0.847	0.068	0.095	0.858	0.153	0.748	0.955
F test (1^{st} stage)	32.329	35.134	0.034	34.064	36.742	0.017	35.158	37.464	0.002
Panel C: Weibull MLE - ITT									
Below mean	-0.017	-0.022	0.014	-0.020	-0.025	0.013	0.016	-0.024	0.052
	(0.0218)	(0.0229)	(0.0448)	(0.0210)	(0.0220)	(0.0439)	(0.0305)	(0.0395)	(0.0486)
Above mean	-0.009	-0.015	0.025	-0.005	-0.011	0.030	-0.016	-0.018	0.014
	(0.0213)	(0.0222)	(0.0437)	(0.0210)	(0.0218)	(0.0439)	(0.0205)	(0.0214)	(0.0433)
N Obs	4664	4492	4664	4664	4492	4664	4664	4492	4664
p-value equality	0.588	0.621	0.449	0.111	0.166	0.091	0.221	0.858	0.181
Panel D: Weibull MLE - Control Function									
Below mean	-0.125	-0.160	-3.214	-0.102	-0.141	-5.096	-0.114	-0.159	-8.029
	(0.1900)	(0.1784)	(6.4094)	(0.1875)	(0.1768)	(9.1835)	(0.1893)	(0.1827)	(24.7565)
Above mean	-0.119	-0.156	-3.231	-0.087	-0.127	-4.997	-0.142	-0.145	-7.841
	(0.1926)	(0.1811)	(6.4638)	(0.1858)	(0.1751)	(9.0331)	(0.1827)	(0.1737)	(24.0515)
N Obs	4664	4492	4664	4664	4492	4664	4664	4492	4664
p-value equality	0.669	0.732	0.770	0.090	0.123	0.510	0.319	0.713	0.790
F test (1^{st} stage)	32.329	35.134	0.034	34.064	36.742	0.017	35.158	37.464	0.002

Note : This table reports various estimates of equation 1.6. Panel A and B reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. Panel C and D reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin, year of arrival dummies and age at migration, columns (3) add household head characteristics at the time of assignment, (4) add the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. In the first three columns concentration is interacted with a dummy being above/below the average age of arrival in the sample. In the following three columns, it is interacted with below/above average length of stay in COA centres and in the last three with below/above average number of asylum seekers registered in the centres at assignment. Panel B reports 2SLS and panel D a control function approach where the interactions of concentration at age 15 are instrumented with the interactions of concentration at assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. F test refers to the refers to the Cragg Donald minimum eigenvalue test. P-value refers to the p-value of the null equality above and below coefficients. ǂǂis used when rounded estimates are < 0.0

Table A12: Alternative IV estimation - Immigrants from the same origin

	(1)	(2)	(3)	(4)	(5)
Ethnic concentration	-2.887	-10.308	121.530	-1.294	-3.442
	(18.7753)	(229.2515)	(32372.6663)	(5.3940)	(39.5997)
Age at migration		-0.357	4.640	-0.057	-0.140
		(7.8006)	(1237.3846)	(0.2163)	(1.5432)
N Obs	4909	4909	4909	4717	4909
Mean Outcome	0.427	0.427	0.427	0.424	0.427
F test	0.023	0.002	0.000	0.058	0.007

Note : The table reports 2SLS estimates of equation 1.4 for the outcome having taken contraceptives at least once by the age of 20. All specifications control for country of origin and year of arrival dummies, columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. The endogenous variable is log of the number of immigrants from the same origin (first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the log of the number of asylum seekers who have been assigned to the same zip4 between 1996 and the year of assignment. The F test is the test statistics for zero effect of the instrument in the first stage. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table A13: Results Municipality Level - Naive Estimates - Immigrants from non Western countries

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	ǂǂ	ǂǂ	-0.001	-0.008	ǂǂ	-0.002	-0.003	-0.005	-0.005	0.913
	(0.0046)	(0.0047)	(0.0047)	(0.0051)	(0.2004)	(0.0156)	(0.0157)	(0.0158)	(0.0175)	(0.6889)
Age at migration		-0.005	-0.004	-0.005	0.002		-0.005	ǂǂ	-0.001	0.023
		(0.0038)	(0.0040)	(0.0040)	(0.0059)		(0.0129)	(0.0136)	(0.0137)	(0.0211)
N Obs	4659	4659	4659	4649	4659	4436	4436	4436	4427	4436
Mean Outcome	0.425	0.425	0.425	0.425	0.425	0.396	0.396	0.396	0.396	0.396
R squared	0.053	0.054	0.059	0.058	0.143					
	Panel B: Quadratic specification									
Ethnic concentration	-0.036	-0.039	-0.041	-0.041	1.168*	-0.154	-0.157	-0.157	-0.158	5.307**
	(0.0424)	(0.0424)	(0.0421)	(0.0422)	(0.6930)	(0.1366)	(0.1370)	(0.1366)	(0.1371)	(2.5468)
Ethnic concentration - sq	0.002	0.002	0.002	0.002	-0.057	0.008	0.008	0.008	0.008	-0.273*
Age at migration		-0.005	-0.005	-0.005	ǂǂ		-0.006	-0.001	-0.002	0.012
		(0.0038)	(0.0040)	(0.0040)	(0.0061)		(0.0130)	(0.0137)	(0.0137)	(0.0219)
N Obs	4659	4659	4659	4649	4659	4436	4436	4436	4427	4436
Mean Outcome	0.425	0.425	0.425	0.425	0.425	0.396	0.396	0.396	0.396	0.396
R squared	0.054	0.054	0.059	0.059	0.143					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the municipality as of January 1^{st} of the year where women turn 15. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at age 15) level. ǂǂis used when rounded estimates are < 0.00

Table A14: Results Municipality Level - ITT Estimates - Immigrants from non Western countries

	Linear Regressions					Weibull MLE				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.005	-0.005	-0.005	-0.002	0.002	-0.020	-0.020	-0.019	-0.011	-0.032
	(0.0049)	(0.0049)	(0.0049)	(0.0050)	(0.0884)	(0.0170)	(0.0170)	(0.0171)	(0.0174)	(0.3157)
Age at migration		-0.004	-0.003	-0.003	-0.004		0.001	0.005	0.004	0.003
		(0.0039)	(0.0041)	(0.0042)	(0.0042)		(0.0131)	(0.0139)	(0.0143)	(0.0146)
N Obs	4499	4499	4499	4307	4499	4279	4279	4279	4099	4279
Mean Outcome	0.425	0.425	0.425	0.423	0.425	0.396	0.396	0.396	0.396	0.396
R squared	0.057	0.057	0.062	0.063	0.113					
	Panel B: Quadratic specification									
Ethnic concentration	0.004	0.004	0.004	ŧŧ	0.002	0.095	0.095	0.076	0.037	-1.587
	(0.0098)	(0.0098)	(0.0098)	(0.0099)	(0.0262)	(0.1246)	(0.1247)	(0.1251)	(0.1302)	(1.0978)
Ethnic concentration - sq	-0.001	-0.001	-0.001	ŧŧ	-0.001	-0.007	-0.007	-0.006	-0.003	0.127
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0022)	(0.0072)	(0.0072)	(0.0072)	(0.0075)	(0.0885)
Age at migration		-0.006	-0.005	-0.005	-0.006		0.001	0.005	0.004	0.003
		(0.0038)	(0.0040)	(0.0041)	(0.0040)		(0.0131)	(0.0139)	(0.0143)	(0.0146)
N Obs	4909	4909	4909	4717	4909	4279	4279	4279	4099	4279
Mean Outcome	0.427	0.427	0.427	0.424	0.427	0.396	0.396	0.396	0.396	0.396
R squared	0.055	0.055	0.060	0.061	0.110					
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports linear regressions for the outcome having taken contraceptives at least once by the age of 20. The right part reports Weibull maximum likelihood estimations where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the municipality as of January 1^{st} of the year of assignment. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ŧŧis used when rounded estimates are < 0.00

Table A15: Results Municipality Level - IV Estimates - Immigrants from non Western countries

	Linear Regressions					(Weibull MLE)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.015	-0.015	-0.015	-0.003	-0.125	-0.066	-0.066	-0.064	-0.030	-0.377
	(0.0228)	(0.0227)	(0.0237)	(0.0235)	(1.0804)	(0.0805)	(0.0803)	(0.0836)	(0.0837)	(3.7351)
Age at migration		-0.004	-0.003	-0.002	-0.009		††	0.005	0.006	-0.011
		(0.0041)	(0.0044)	(0.0044)	(0.0545)		(0.0138)	(0.0148)	(0.0152)	(0.1887)
N Obs	4314	4314	4314	4131	4314	4106	4106	4106	3934	4106
Mean Outcome	0.423	0.423	0.423	0.421	0.423	0.394	0.394	0.394	0.392	0.394
R squared	0.054	0.054	0.059	0.063						
F test (1^{st} stage)	118.930	119.592	109.882	110.472	0.087	118.930	119.592	109.882	110.472	0.087
	Panel B: Quadratic specification									
Ethnic concentration	0.225	0.222	0.169	0.167	-2.081	0.650	0.650	0.509	0.350	-8.147
	(0.2826)	(0.2842)	(0.2864)	(0.3046)	(8.1337)	(0.9829)	(0.9865)	(1.0094)	(1.1018)	(13.0008)
Ethnic concentration - sq	-0.013	-0.013	-0.010	-0.009	0.153	-0.039	-0.039	-0.031	-0.021	0.665
	(0.0151)	(0.0151)	(0.0152)	(0.0162)	(0.4985)	(0.0525)	(0.0527)	(0.0537)	(0.0588)	(0.8036)
Age at migration		-0.003	-0.002	-0.001	0.024		0.003	0.007	0.008	0.168
		(0.0044)	(0.0046)	(0.0048)	(0.0944)		(0.0147)	(0.0158)	(0.0164)	(0.1557)
N Obs	4314	4314	4314	4131	4314	4106	4106	4106	3934	4106
Mean Outcome	0.423	0.423	0.423	0.421	0.423	0.394	0.394	0.394	0.392	0.394
R squared	0.044	0.044	0.053	0.057						
F test (1^{st} stage)	47.809	47.582	44.776	38.632	0.074	47.809	47.582	44.776	38.632	0.074
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The left part of the table reports 2SLS estimates for the outcome having taken contraceptives at least once by the age of 20. The right part reports control function estimates from duration analysis where failure is defined as the age between 15 and 20 when a woman starts taking contraceptives. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of non-Western immigrants(first and second generations) living in the municipality as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. The sample only includes complete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level.

Table A16: Falsification Test - Educational Outcomes - Naive Estimates - Immigrants from non Western countries

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	0.007	0.009	0.009	0.012	0.024**	0.008	0.006	0.007	0.007	0.019**
	(0.0070)	(0.0069)	(0.0067)	(0.0072)	(0.0115)	(0.0055)	(0.0056)	(0.0053)	(0.0060)	(0.0094)
Age at migration		0.010**	0.010**	0.009**	0.012**		-0.005**	-0.006**	-0.006**	-0.005**
		(0.0040)	(0.0041)	(0.0041)	(0.0042)		(0.0023)	(0.0023)	(0.0023)	(0.0024)
N Obs	4592	4592	4592	4581	4592	6955	6955	6955	6921	6955
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.158	0.244	0.092	0.092	0.136	0.136	0.193
	Panel B: Quadratic specification									
Ethnic concentration	0.095**	0.095**	0.076*	0.065	0.053	0.048	0.049	0.029	0.020	0.008
	(0.0481)	(0.0480)	(0.0449)	(0.0466)	(0.0535)	(0.0406)	(0.0407)	(0.0377)	(0.0398)	(0.0458)
Ethnic concentration- sq	-0.007*	-0.007*	-0.005	-0.004	-0.002	-0.003	-0.003	-0.002	-0.001	0.001
	(0.0037)	(0.0036)	(0.0034)	(0.0035)	(0.0043)	(0.0031)	(0.0031)	(0.0029)	(0.0031)	(0.0037)
Age at migration		0.010**	0.010**	0.009**	0.012**		-0.005**	-0.006**	-0.006**	-0.005**
		(0.0040)	(0.0041)	(0.0041)	(0.0042)		(0.0022)	(0.0023)	(0.0023)	(0.0024)
N Obs	4592	4592	4592	4581	4592	6955	6955	6955	6921	6955
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.119	0.120	0.159	0.158	0.244	0.092	0.093	0.136	0.136	0.193
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the age of 15. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year where women turn 15. Standard errors are clustered at the year of arrival/municipality (at age 15) level.

Table A17: Falsification Test - Educational Outcomes - ITT Estimates - Immigrants from non Western countries

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.002	-0.002	-0.002	-0.002	-0.002	-0.001	-0.001	††	-0.001	-0.001
	(0.0060)	(0.0059)	(0.0057)	(0.0064)	(0.0111)	(0.0049)	(0.0049)	(0.0047)	(0.0048)	(0.0092)
Age at migration		0.010**	0.010**	0.009**	0.011**		-0.005**	-0.006**	-0.006**	-0.006**
		(0.0038)	(0.0038)	(0.0039)	(0.0039)		(0.0022)	(0.0023)	(0.0024)	(0.0024)
N Obs	4593	4593	4593	4421	4593	7020	7020	7020	6754	7020
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.119	0.158	0.161	0.203	0.092	0.092	0.136	0.138	0.165
	Panel B: Quadratic specification									
Ethnic concentration	-0.055*	-0.054*	-0.052*	-0.065*	-0.060	-0.028	-0.028	-0.028	-0.043	-0.009
	(0.0298)	(0.0297)	(0.0286)	(0.0384)	(0.0445)	(0.0256)	(0.0257)	(0.0244)	(0.0306)	(0.0367)
Ethnic concentration- sq	0.005*	0.005*	0.004*	0.006	0.006	0.002	0.002	0.002	0.004	0.001
	(0.0027)	(0.0027)	(0.0025)	(0.0034)	(0.0044)	(0.0023)	(0.0023)	(0.0021)	(0.0027)	(0.0036)
Age at migration		0.010**	0.010**	0.009**	0.011**		-0.005**	-0.006**	-0.006**	-0.006**
		(0.0038)	(0.0038)	(0.0039)	(0.0039)		(0.0022)	(0.0023)	(0.0024)	(0.0024)
N Obs	4593	4593	4593	4421	4593	7020	7020	7020	6754	7020
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.118	0.120	0.159	0.162	0.203	0.092	0.092	0.136	0.138	0.165
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics when the eldest daughter in the sample turn 15 y.o., (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at the time of assignment. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ††is used when rounded estimates are < 0.00

Table A18: Falsification Test - Educational Outcomes - IV Estimates - Immigrants from non Western countries

	Smaller Sample					Larger sample				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
	Panel A: Linear specification									
Ethnic concentration	-0.019	-0.019	-0.020	-0.018	0.352	-0.005	-0.006	++	-0.009	0.161
	(0.0532)	(0.0531)	(0.0536)	(0.0534)	-24599	(0.0326)	(0.0331)	(0.0323)	(0.0299)	-29590
Age at migration		0.009**	0.009*	0.008*	0.029		-0.005**	-0.006**	-0.006**	0.002
		(0.0045)	(0.0047)	(0.0048)	(0.1232)		(0.0027)	(0.0028)	(0.0028)	(0.1470)
N Obs	4593	4593	4593	4421	4593	7020	7020	7020	6754	7020
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared	0.115	0.116	0.155	0.158		0.091	0.091	0.136	0.136	0.029
F test (1^{st} stage)	39.934	39.405	36.182	43.674	0.040	95.016	91.716	87.252	100.021	0.015
	Panel B: Quadratic specification									
Ethnic concentration	-1.631	-1.604	-1.434	-1.399	0.558	-0.514	-0.525	-0.450	-0.652	-0.672
	(1.0714)	(1.0653)	(0.9553)	(1.0087)	(2.7261)	(0.5440)	(0.5439)	(0.4886)	(0.5447)	(37.9590)
Ethnic concentration- sq	0.125	0.123	0.110	0.109	-0.008	0.042	0.042	0.037	0.054	0.043
	(0.0809)	(0.0804)	(0.0721)	(0.0768)	(0.2014)	(0.0430)	(0.0430)	(0.0385)	(0.0438)	(21152)
Age at migration		0.009*	0.010**	0.010*	0.034		-0.004	-0.004	-0.003	-0.011
		(0.0052)	(0.0051)	(0.0052)	(0.0254)		(0.0029)	(0.0030)	(0.0034)	(0.5155)
N Obs	4953	4953	4953	4421	4953	7020	7020	7020	6754	7020
Mean Outcome	0.552	0.552	0.552	0.553	0.552	0.564	0.564	0.564	0.565	0.564
R squared								0.053		
F test (1^{st} stage)	3.369	3.347	3.835	4.301	0.748	4.500	4.543	5.158	4.856	++
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year of arrival FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parental Characteristics	✗	✗	✓	✓	✓	✗	✗	✓	✓	✓
Municipality FE	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓
Neighbourhood index	✗	✗	✗	✓	✗	✗	✗	✗	✓	✗

Note : The upper part of this table reports estimations from equation 1.4 and the lower part from equation 1.5. The outcome variable is having at least attended HAVO as highest education level of education. The left part of the table reports results on the sample of complete spells, the right part of the table on complete and incomplete spells. All specifications control for country of origin and year of arrival dummies, default results in columns labelled (1). Additional controls are added successively, (2) adds age at migration, (3) adds household head characteristics at the time of assignment, (4) adds the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. In panel A, the endogenous variable is log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year when the young woman turns 15. The instrument is the same measure the year of assignment. In panel B, the two endogenous variables are concentration and its square at age 15 and the two instruments are concentration and its square at assignment. The F test in panel A test for zero effect of the instrument in the first stage, while in panel B it refers to the Cragg Donald minimum eigenvalue test. Standard errors are clustered at the year of arrival/municipality (at assignment) level. ++is used when rounded estimates are < 0.00

Table A19: Falsification test - Heterogeneity - Immigrants from non Western countries

	Age at arrival			Length of Stay			Centre Size		
	(3)	(4)	(5)	(3)	(4)	(5)	(3)	(4)	(5)
Panel A: OLS - ITT - Complete Spells									
Below mean	-0.001	-0.001	-0.001	-0.002	-0.002	-0.003	-0.008	-0.008	-0.006
	(0.0061)	(0.0068)	(0.0114)	(0.0061)	(0.0067)	(0.0112)	(0.0072)	(0.0084)	(0.0130)
Above mean	-0.003	-0.003	-0.003	-0.002	-0.003	-0.002	-0.001	-0.002	-0.001
	(0.0060)	(0.0067)	(0.0112)	(0.0058)	(0.0065)	(0.0112)	(0.0058)	(0.0065)	(0.0112)
p-value	0.690	0.583	0.685	0.974	0.775	0.740	0.254	0.328	0.538
N obs	4593	4421	4593	4593	4421	4593	4593	4421	4593
Panel B: 2SLS - Complete Spells									
Below mean	-0.020	-0.018	0.353	-0.020	-0.019	0.500	-0.020	-0.021	-2.040
	(0.0535)	(0.0532)	-25085	(0.0526)	(0.0526)	-45448	(0.0534)	(0.0541)	(94.5063)
Above mean	-0.021	-0.020	0.353	-0.020	-0.020	0.493	-0.013	-0.013	-1.955
	(0.0542)	(0.0540)	-25215	(0.0520)	(0.0521)	-44746	(0.0516)	(0.0519)	(90.8066)
p-value	0.657	0.552	0.988	0.965	0.828	0.923	0.213	0.280	0.982
N obs	4593	4421	4593	4593	4421	4593	4593	4421	4593
F test	29.954	32.745	0.027	31.269	33.962	0.013	32.791	35.110	ŧŧ
Panel C: OLS - ITT - Complete and incomplete Spells									
Below mean	0.001	-0.001	ŧŧ	0.001	-0.001	ŧŧ	-0.006	-0.010	-0.002
	(0.0052)	(0.0052)	(0.0096)	(0.0050)	(0.0051)	(0.0093)	(0.0064)	(0.0072)	(0.0106)
Above mean	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	0.001	-0.001	-0.000
	(0.0049)	(0.0050)	(0.0091)	(0.0047)	(0.0048)	(0.0092)	(0.0047)	(0.0048)	(0.0092)
p-value	0.587	0.622	0.668	0.665	0.606	0.702	0.181	0.147	0.753
N obs	7020	6754	7020	7020	6754	7020	7020	6754	7020
Panel D: 2SLS - Complete and incomplete Spells									
Below mean	ŧŧ	-0.009	0.150	-0.001	-0.010	0.390	-0.004	-0.016	-0.072
	(0.0321)	(0.0297)	-25112	(0.0320)	(0.0297)	(7.5493)	(0.0329)	(0.0310)	(3.3698)
Above mean	-0.001	-0.011	0.151	-0.002	-0.011	0.385	0.003	-0.006	-0.067
	(0.0326)	(0.0304)	-25574	(0.0317)	(0.0295)	(7.4735)	(0.0316)	(0.0293)	(3.2341)
p-value	0.585	0.579	0.979	0.667	0.618	0.949	0.192	0.130	0.970
N obs	7020	6754	7020	7020	6754	7020	7020	6754	7020
F test	68.330	74.327	0.012	69.442	75.234	0.003	71.913	77.761	0.007

Note : This table reports various estimates of equation 1.6. The outcome variable is having at least attended HAVO as highest education level of education. All specifications control for country of origin, year of arrival dummies and age at migration, columns (3) add household head characteristics at the time of assignment, (4) add the neighbourhood quality index while (5) add municipality fixed effects. Both geographical controls in (4) and (5) refer to where women live at assignment. Ethnic concentration is measured as the log of the number of non-Western immigrants (first and second generations) living in the zip 4 as of January 1^{st} of the year of assignment. In the first three columns concentration is interacted with a dummy being above/below the average age of arrival in the sample. In the following three columns, it s interacted with below/above average length of stay in COA centres and in the last three with below/above average number of asylum seekers registered in the centres at assignment. Panel B and D report 2SLS where the interactions of concentration at age 15 are instrumented with the interactions of concentration at assignment. Panel A and B include complete spells, panel C and D complete and incomplete spells. Standard errors are clustered at the year of arrival/municipality (at assignment) level. F test refers to the refers to the Cragg Donald minimum eigenvalue test.P-value refers to the p-value of the null equality above and below coefficients. ŧŧis used when rounded estimates are < 0.0

Chapter 2

Who influences young immigrants?

with Eva Johansen

1 Introduction

What determines the adoption (or rejection) of natives culture by young immigrants? Does exposure to young natives foster the adoption of mainstream behaviours? Are young immigrants responsive to other young natives, in the sense that they adapt their own behaviours to natives? Or are they closed to their influence but opened to that of other immigrants instead?

It is difficult to provide convincing empirical evidence on those questions for three reasons. First, one needs to isolate a behaviour for which natives and immigrants differ and for which adoption by immigrants is non-trivial. Once this outcome is selected, one needs to provide exogenous variation in exposure to that behaviour. People choose the neighbourhoods in which they live, the schools to which they go and the network of friends to which they belong, so one should be careful not to mistake sorting for peer influence. Third, once random variation in exposure is found, one must find a way to address the reflection problem, Manski (1993). If teenagers are in contact with each other, how can we be sure to isolate a one-sided, from a reciprocal, influence? This paper addresses these three challenges.

We focus on teenage immigrants decision to use contraceptives. There are particular wide gaps in culture between Western and non-Western countries when it comes to teenage female sexuality[1].

[1]In a descriptive study on second generation immigrants in Canada, Lalonde and Giguere (2008) singles out attitudes towards pre-marital sex as one where immigrants from non Western origins feel torn between their cultural heritage and the mainstream views in the country. "Of greater interest, however, were ratings about the perceived appropriateness of engaging in premarital sex if involved in a loving relationship. As expected, South Asian and Chinese Canadians perceived premarital sexual intercourse as less appropriate than did their Canadian peers, although this difference was more marked for South Asian Canadians. More importantly, the ratings of South Asian Canadians fell in between their South Asian parent perceived as appropriate (i.e., not appropriate) and what they perceived their Canadian peers perceived as appropriate (i.e., quite appropriate). These data provide evidence that second generation South Asians see their views regarding sexuality as falling between two sets of cultural norms."

This is a particularly important area as female sexuality is potentially connected to gender role and is the predecessor for fertility and marriage - both factors that are important for later labour market attachment Goldin and Katz (2002). In this paper, we rely on administrative data on contraceptive usage, abortion, treatment for a STD and visit to a GP[2]. All the data comes from administrative registries and are not subject to the measurement error problems that could arise from using survey data.

For instance, 45.54% of Danish women have used contraceptives at least once by the age of 16 years old when only 10.09% of teenagers with a non-Western background (first or second generation immigrants) have done so. At age 20, the numbers rise to 85.73% for natives and 40.81% for non-Western immigrants.

It is important to note that we do not focus on self-reported identity but on information about a behaviour. Identity and behaviour are two different things and evidence on one is not necessarily indicative of the other. In this paper, we prefer to look at behaviour for two reasons, (i) it captures preferences better than attitudes do and (ii) identity among young immigrants can have multiple layers. Attachment to a foreign heritage can co-exist with behaving like a native.

To address selection, we rely on cross cohort variation on schoolmates behaviour in a fashion initiated by Hoxby (2000). High school is a central place of teenagers' social life and as such a natural starting point to look at peer effects. We control for sorting into high schools by using a large set of school fixed effects and school-specific time trend. This isolates a residual variation that we argue is random. There is a difference (at the cohort level) of 0 to 4 girls between the actual number of non-Western immigrants who already took contraceptives and the number predicted by various sets of fixed effects. This is the variation we use for identification.

A key element to credibly identify peer effects, Angrist (2014), is to clearly establish which group is influencing, which group is influenced and to make sure that influence goes only in one direction. In this paper, we look at the influence of older cohorts on younger ones, assuming that younger girls are influenced by the sexual behaviour of older girls and not the other way around[3].

To recap in broad (and simplified) terms the empirical strategy: in a given school, there are certain years where slightly more/less girls from older cohorts take contraceptives than usual. If that variation induces immigrant women from younger cohorts to take contraceptives (or not), we consider it as evidence of peer effects. Whether immigrant teenagers are responsive to the behaviour of natives or of other immigrants is indicative of which group is influential in the process of cultural adaptation.

The main result of this paper is that older immigrant women have an influence on younger ones while older native women do not. An increase in contraceptive usage by non-Western immigrants

[2]As explained more in detail below, in Denmark if a teenager wants to see a GP with the specific purpose of discussing contraception, this appointment appears specifically in the data.

[3]This assumption is similar to Clark and Loheac (2007) who looks at schoolmates influence on risky behaviour and to Altonji et al. (2017) who looks at siblings influence, both assuming older influences younger and not vice versa.

in the second and third years of high school, i.e. the last years, leads to a decrease in usage by first year students of the same group. This effect is long lasting since probability of using contraceptives remain lower three years later. This effect does not appear when there is a change in the behaviour of native teenagers.

While the direction of the influence (positive or negative) is not the main focus of this paper, the fact that the main result is negative is interesting and intriguing by itself. To give more strength to this result, we reproduce the analysis on peer influence with native teenagers. We find that teenagers who do not have a foreign background also have a lower probability of using contraceptives when more of their native peers do so. Interestingly, first year native high-school girls respond to changes in older immigrants girls behaviour (although less strongly than to natives).

To better understand the mechanisms at play, we look at whether the change in contraceptive behaviour is explained by a change in sexual behaviour or a change in medication alone. To do so, we see if the variation in peers usage of contraceptives also influence the probability of having an abortion and the probability of being treated for chlamydia, the most common STD among young Danes. Likelihood of having had an abortion or have been treated for chlamydia by the end of high school is very low (respectively 3.27% and 3.05% probability for non-Western immigrants). Thus, it is harder to pick up an effect that is statistically significant than when looking at contraceptive usage. Although results are less precise, they clearly point to a negative effect on abortions and a null on chlamydia. We take it as evidence that changes in contraception goes through a change in sexual behaviour.

The rest of the paper is organized as follows: section 2 reviews the relevant literatures and how this paper contributes to them. Section 3 details the institutional settings and describes the data. Section 4 presents descriptive evidence while section 5 explains the identification strategy and shows balancing tests. Baseline results on peer effects are in section 6, while section 7 provides additional results on mechanisms. Section 8 shows robustness checks and section 9 concludes.

2 Literature Review and Contribution

Literature on cultural assimilation The main literature this paper contributes to is the one on the determinants of cultural assimilation. Bisin et al. (2016) have established theoretically under which conditions immigrants form an identity characterised by *conformity* or *distinction*. Their empirical evidence points towards identity being formed as a cultural distinction mechanism. Olcina et al. (2017) model how individuals balance agreeing with their personal ideas and assimilate to average peers, the sort of trade-off we analyse empirically.

The role played by friends from school on cultural integration has been studied in Patacchini and Zenou (2016). The authors establish that vertical integration, defined as the decision by parents to transmit their cultural heritage to their children, is complementary with horizontal integration, defined as the influence of friends in schools. We focus here on the role played by

various peer groups in the horizontal integration.

This paper also contributes to the discussion on who influences the behaviours of immigrants. Theoretical contributions, from network theory Verdier and Zenou (2017), evolutionary game theory Kuran and Sandholm (2008); Olcina et al. (2017) and the literature on cultural leaders Verdier and Zenou (2018) have studied the dynamics of preferences and behaviours. In Kuran and Sandholm (2008); Olcina et al. (2017), a minority individual is paired with another agent and has to decide whether or not to adjust her behaviour. Verdier and Zenou (2018) focuses on the role of cultural leaders (influential institutions in their communities) on the dynamics of cultural integration. Olcina et al. (2017) looks at the role of social networks structure on the dynamics of cultural assimilation. This paper provides empirical evidence to complement these theoretical works.

In relation to models using evolutionary game theory, this paper looks at whether immigrants would try to conform with both other immigrants and natives when paired with them. In relation to the models derived from network theory, we investigate if both natives and immigrants can have an influential role. The approach is methodologically different, in the sense that we do not model the network. We rely instead on an exogenous shock to the behaviour of several distinct groups and see who the population of interest is responsive to.

Literature on peer effects This paper also contributes to the vast literature on peer effects. The objective of studies in this literature is to develop an empirical strategy to assess if an influential group influences an influenced group about a specific behaviour. Interactions at the school are a natural candidate for such studies. Angrist and Lang (2004); Ballatore et al. (2018) have for instance respectively looked at the effects of minority students in the US and immigrant students in Italy on the academic performances of non-minority/natives. While results appear small and short lived in the US, they are larger (and negative) in the case of Italy.

A popular methodology has been to rely on cross cohort variations in peer characteristics. The idea is to control for sorting into schools by using a large set of fixed effects, namely school FE, cohort FE and eventually school specific time trend. This allows to see which peers characteristics influences performance at school Hoxby (2000); Black et al. (2013), chosing a STEM major at university Brenoe and Zolitz (2018) or female labour supply decision at adulthood Olivetti et al. (forthcoming). We adopt this strategy with a little twist. To avoid the reflection problem, Manski (1993), and follow the guideline established in Angrist (2014), we clearly define a group that is influential but is not influenced in return. Following Clark and Loheac (2007); Altonji et al. (2017), this group is composed of older students in the same school.

Risky behaviours in the teenage population has been a particular topic of interest. Argys and Rees (2008) studies the effect of having older schoolmates in your cohort. They rely on variation in the age at which other students start kindergarten. They find a positive effect on the probability for young women to use illicit substance, so does Altonji et al. (2017) when focusing on older siblings rather than schoolmates. Clark and Loheac (2007) look at substance use and the effect

of lagged peers behaviour, from older cohorts. We use a similar structure of peers influence in this paper. Card and Giuliano (2013) look in particular at sexual initiation of teenagers and find that peers have a very strong effect, in particular for young women. Many of these papers point to older schoolmates having influence on younger peers and, to influence them to engage in risky behaviours. This gives support to our strategy of using older schoolmates as role models. Our results however, are in contrast with theirs, as we find a negative effect on sexual behaviour.

The closest paper to ours is Merlino et al. (forthcoming), where the authors use the same identification strategy to investigate a (relatively) similar question. The objective of this paper is to see if an increase in exposure to peers from a different racial background in childhood increases the likelihood of being married to someone from that ethnicity later in life. Using the wave 4 of the Add dataset, they find positive effects.

3 Institutional Settings and Data

3.1 Information on schools

After grade 9 (the end of compulsory education), students can choose to enter academic high school or a vocational program. This paper focuses on women, who enter one of the four major types of academic high school: general (STX), business (HTX), two year general (HF) or technical (HHX)[4]. Before 2005 the general high school was divided into a language and a math track. After 2005 the two tracks are combined and students choose subjects more freely (thus, the cohort size increases from 2005). A school is defined as one type of high school in one school[5].

High school takes three years (except from the two year general high school) and the average age for starting high school is 16. High school represents a change in the peer environment because it implies a new school with teenagers from different lower secondary schools. Thus, new and older cohorts is a mix of teenagers from a woman's own lower secondary school and from other nearby lower secondary schools. Table 1 shows the distribution of school types (both at the program and student level) together with information on the share non-Western immigrants.

Table 1 here

We follow 213,346 students (natives and first/second generation immigrants) in 643 schools. Most of them (79.65%) are in general (STX) programs and relatively few are in HF and HTX. On average in the entire sample, 6.04% of students are non-Western immigrants. This share is lowest in HF (2 years general) programs, 5.55%, and highest in STX (combined) programs, 7.17%. There

[4]At the end of the 9^{th} grade, students can also decide to take an optional 10^{th} grade and postpone their entry into high school. We focus on the year of entry in high school after 9^{th} or 10^{th} grade.

[5]Different types of high school can be organized in the same building, but they are still referred to as different schools. This is mostly the case for the language and math track in the general high school.

appears to be an increase in the share of non-Western immigrants over time, the proportion in combined STX programs after 2005 is larger than the two separate components of the program (Language and Math) before 2005.

Table 2 here

Table 2 breaks down the sample by origin. Panel A shows how many women are from a Western versus a non-Western background. Obviously the largest share of teenagers attending Danish schools is made of natives and immigrants from a Western background, 94.34%. Non-Western immigrants represent 5.66% of all women attending high school or 12,082 people. Panel B breaks down whether high school students have origins from a mostly Muslim country as it was singled out in the literature, Bisin et al. (2008), as being a characteristic affecting immigrants attitudes towards mixing with natives[6].

It is important to emphasize that we do not have information on individual religious affiliation, neither on religious practises. We just match individuals with the faith mainly practised in their origin country. 7,718 out of the 12,082 (64%) non-Western immigrants come from a Muslim country. The three largest communities among non-Western immigrants are Turkish, Pakistani and people from Bosnia-Herzegovina. They each account for more than 1,000 individuals in the sample.

3.2 Information on contraceptives and other health measures

We use five different health measures in this paper: usage of contraceptives, abortion, treatment of chlamydia, information on whether a young woman saw her GP to talk about contraception and whether she saw a gynaecologist. All this information comes from administrative registries.

Our measure of contraception includes the pill (the vast majority), rings, patches, injections, IUDs (intrauterine device) and implants. To be more precise, it includes the following entries of the ATC classification (for drugs), G03AA, G03AB, G03AC, G02BB01, G02BA and information on doctors inserting an IUD. We have data on drugs collected, not on drugs prescribed. If a drug is prescribed but not bought, it does not appear in the data. We adopt a relatively large definition of contraceptive, instead of only looking at the pill (by far the most common mean of contraception at these ages), to use all the information available to us.

Our measure does not account for emergency contraceptives. Contraceptives in Denmark are accessible to young women. They can be prescribed by a GP and not necessarily by a gynaecologist. The legal age to be prescribed contraceptives without parental approval is 18 years old. However, doctors usually prescribe them without informing parents before that age. Consultation with a GP

[6]We characterize a country as mostly Muslim, if it is a member of the Organisation of Islamic Cooperation. Member states are listed in table A1.

(and therefore the prescription) is free. Young women still have to pay for the product themselves. Prices are around 10 euros for three months of the pill.

All abortions performed in Denmark are reported in administrative registries. They are free of charge. Parental consent is necessary for abortions on teenagers below 18 years old. We have information on STDs, in particular on having been treated for chlamydia. It is the most widespread sexual disease among young Danes. It can be measured by looking at the prescription of a particular antibiotics in a particular dosage[7]. Many people do not have any symptoms of this disease, although they suffer from it. This means that our measure of chlamydia should be interpreted as having being diagnosed with chlamydia, so both having been positively tested for it or having seen a doctor following the apparition of symptoms, rather than simply having chlamydia.

Medical registries also include two informations on doctors appointments: whether a woman has had a consultation with a gynaecologist and whether a woman had an appointment with her GP specifically to talk about contraception. In Denmark, the state pays for the GP. There are fixed payments for different services. One is a standard payment for a consultation, but there is an additional amount when the consultation is about contraception. This ensures that these appointments are well reported. In the rest of the paper, consultation to a GP refers to consultation about contraception.

3.3 Additional data

There are two additional types of data used in this paper: information on (i) socio-economic conditions and family situation of young women and (ii) on their origins. The former type of variables are used to perform balancing tests and is added as controls in the main regressions. These characteristics include parents marital status, parents employment status, parents education level, parents age at birth (of their daughter) and age at which a young woman starts high school. The background characteristics are measured at age 13, before the measurement of the explanatory variables of interest.

Administrative registries contain information on country of origin for both first and second generation immigrants. We use two categories of origin, non-Western immigrants (relying on the classification established by Statistics Denmark) as opposed to Western immigrants (both native Danes and immigrants with a Western background) and Muslim origin (based on being from a country that is part of the Organisation for Islamic Cooperation) as opposed to non-Muslim origin. In what follows, *natives* refers to both natives and people without a non-Western background and *immigrants* refer to (1^{st} and 2^{nd} generation) individuals with a non-Western background.

[7] ATC code J01FA10 with the strength 500 mg and pack size 2. The guidelines say that 1,000 mg from two pills taken right after each other is the recommended way to treat chlamydia.

3.4 Sample definition

The sample consists of women born between 1981 and 1993 who enrol for the first time in high school between the year they turn 14 and 19 and who live in Denmark as of January,1^{st} of the year they turn 13 years old. Since data on drugs is only available from 1995 to 2015, 1981 becomes automatically the lower limit and 1993 the upper one. We restrict the sample to women who enrol in high school in July through September, which is in the beginning of the school year. We only consider schools with more than 10 students (both boys and girls) in a cohort where 95 percent or more start between the year they turn 14 to 19. We do this to exclude women enrolled in very small schools and schools where most peers are not young women (19 years old being the upper limit to be characterised as young). Table A2 shows how many observations are deleted by imposing each constraint of the sample definition detailed above. These sample restrictions are similar to those imposed by Brenoe and Zolitz (2018).

4 Descriptive Statistics

4.1 Difference between behaviour of natives and immigrants

The purpose of this subsection is to see how each of the five measures differs between immigrants and natives. It justifies the use of these variables as outcomes and introduce the discussion of what exactly is measured with each of them. It also helps understanding the magnitude of the effects we find later in the paper.

Table 3 reports for all ages between 15 and 20 years old the share of natives and immigrants who have taken contraceptives at least once, who ever had an abortion, who were ever treated for chlamydia, who saw a gynaecologist at least once and who had at least one appointment with a GP to specifically talk about contraception. All these measures are cumulative. At the individual level, you can only stay at zero or go from zero to one and, at the aggregate level, it cannot decrease.

Table 3 here

What is striking from this table is that certain behaviours show very strong differences between natives and immigrants while others do not. For instance, there is a clear difference in contraceptive usage. At every age, the share of immigrants who have ever used contraceptives is between twice and five times lower than the share of natives. A similar picture emerges from appointments to the GP. Although the numbers are very close for these two categories, the correlation coefficient is only 0.76, which is high in absolute but lower than what the aggregate numbers would suggest.

There are many reasons why immigrant women who would be willing to take contraceptives do not start using them. The first that comes to mind is that they could fear that their parents

find out. Talking about contraception with a doctor in the secrecy of a medical appointment is a measure more likely to capture willingness of young women. This outcome has limitations since teenagers may internalize the constraints of their environment in their decision to mention contraception with a GP. It is still an interesting complement to the outcome taking contraceptives. From table 3, immigrant women are less incline to mention contraception.

The picture is different for the outcomes having had an abortion and going to see a gynaecologist. For these ones, natives and immigrants are very similar and the prevalence of these behaviours in the general population is much more limited. Teenagers in Denmark do not see a gynaecologist to get a prescription for their contraception. As one needs a referral to see a gynaecologist, women only see one if they have been to a GP before. This outcome typically indicates some kind of disease, which could explain the little difference between natives and immigrants. It however shows that there are no striking differences in the availability of gynaecological health care between young natives and immigrants. Abortion captures unprotected sexual intercourses. There is little difference between the two groups for this measure.

The probability of being treated for chlamydia is also very different between the two groups, more or less three times higher for natives than for immigrants. Although, there is a caveat with this outcome, the spread of STDs is related to having more partners and table 3 points towards immigrants and natives having different sexual behaviours.

Putting together these different elements, using contraceptives for natives is somehow a norm in the sense that almost every young woman uses them. This norm is not shared by immigrants. The latter group seems much less eager to use contraceptives. For these reasons, our main outcome is usage of contraceptives. When administrative data on drug usage is crossed with surveys on sexual behaviour, the correlation between starting to take contraceptives and becoming sexually active is very high (0.9). From the perspective of an economists/econometrician, observing usage of contraceptives in administrative data is a good signal of individual sexual behaviour. Moreover, contraception can also be seen as a woman empowerment device Bailey (2006); Goldin and Katz (2002) and is in itself a meaningful consumption.

Although immigrant teenagers use less contraceptives and that it points towards being less sexually active, a non-negligible share engages in unprotected sex and suffers from STDs. This population can influence other immigrants. Note that the influence could go both ways, other immigrants may feel liberated by observing their peers having more open relations. They can also feel more distant from women with the same background who do not follow the rule of the community. They can also move away from what is perceived as risky behaviours. Having access to additional information on medication, abortions and STDs, allows us to say more about mechanisms. We use the remaining outcomes to understand what is at play behind the effect on usage of contraceptives. Table 4 shows the same information by school grade. The picture is substantially the same.

Table 4 here

4.2 Sample characteristics

In tables 5 and 6, we produce two sets of descriptive statistics about the sample. The first one, compares the mean values of the control variables for young native and young immigrant women. This allows to say how similar/different the two populations are. The second set of descriptive statistics look at the same information in the schools where immigrant women and natives study. This allows to compare the environment in which the two groups live.

Table 5 here

The socio-economic conditions of the two groups are different. Unemployment rate is much higher and education much lower for immigrant fathers, 9% versus 2% for unemployment for immigrants and native fathers and 150 versus 72 months of education for the same groups.

Young immigrants also have different role models when they look at their mothers. The share of immigrant mothers out of the labour force is 48% when it is only 8% for natives. Immigrant mothers are also much less educated than natives, 66 months versus 153. Comparison between the two groups may not give an accurate picture, but it is interesting to observe that native mothers are on average slightly more educated than native fathers (+ 3 months), while immigrant mothers have 6 months less education than fathers. Immigrant parents, both mothers and fathers, are also younger at the birth of their daughter, than natives. Young women, both natives and immigrants start high school at the same age.

When one compares cohort characteristics of the schools where both groups study: immigrant teenagers attend larger schools, 169 vs 155 students, with more immigrants, 15% versus 5% of immigrants girls and 14% versus 6% of immigrant boys. Immigrant peers have less educated parents, 9 months of education less for both mothers and fathers compared to native parents.

Usage of contraceptives among natives and immigrants is the same in the schools in which natives and immigrants go (36% versus 37% for native women of older cohorts in natives and immigrant schools, 7% versus 8% for older immigrant women).

Table 6 here

5 Identification Strategy

5.1 Empirical challenges and how to address them

How to account for sorting into schools? Students sort into schools, therefore any regression on peer characteristics is likely to confound what drove a student to start high school in a particular

place in a particular year and peers influence. Our strategy to account for sorting into schools, consists in approximating the sorting decision with a large set of fixed effects. In particular, we look successively at the inclusion of (i) school fixed effects, (ii) school and time trend (common to everyone), (iii) school effects with a school specific time trend. In the spirit of the Frisch Vaugh Lowell theorem, the residual variation that is not explained by fixed effects is regressed on peers behaviour. Usage of contraceptives can be predicted up to a certain extent by fixed effects. The identifying assumption is that what remains is as good as random.

An alternative way to think about the same identification strategy, is to elicit which groups are being compared to identify peer effects when using each set of regressors. When school fixed effects are used, we compare students from the same school but different cohorts. When school and cohort fixed effects are used, we compare the changes over time between schools, in a fashion similar to a difference in difference. When school and cohort fixed effects are used with a school linear trend, we compare the evolution of changes over time between schools.

How to account for the reflection problem? The objective is to find a group that influences the behaviour of first year high-school girls and which is not influenced back. There is a trade-off between finding a relevant group of peers (with whom there is intense interaction) and not falling in the reflection problem.

Our solution is to focus on older girls (in year 2 and 3) in the same school. This corresponds to the behavioural assumption that younger girls are influenced by the sexual behaviour of older ones and not vice versa. To ensure that influence goes in one direction, we focus on the behaviour of teenagers in older cohorts prior to the new first years joining the school.

A limitation is that interactions with these older girls may be less intense than with teenagers from the same cohort. This is why as a robustness, we focus on girls in the same cohort but only on their behaviour prior to joining high school.

5.2 Evidence on the variation being used

The variation we use is residual, in the sense that we look at peers behaviour conditional on a large set of fixed effects. Thus, is not straightforward to see the magnitude in the variation used to identify peer effects. Compared to what is predicted by the fixed effects, how many more/less peers take contraceptives?

To obtain this information, we do the following: we collapse the information on the number of girls who use contraceptives at the school/cohort level. We do that both for older girls in the same school and girls from the same cohort. We regress this collapsed number on school fixed effects and school specific time trend, the most demanding specification we use later on. The residuals of this regression is the number of girls (who use contraceptives) who were not "anticipated", based on information captured by fixed effects, when one decided to join a certain school a certain year.

We plot the residuals of the predicted value for older immigrant teenagers in figure 1 and older native teenagers in figure 2.

Figure 1 here

Figures 1 and 2 plot the histogram of the residuals together with a fitted normal distribution. When looking at immigrants from older cohorts, we see that most of the residual variation is of the order 0 to 4, meaning that most of the times less than four girls take (or do not take) contraceptives from what could be predicted by fixed effects. In appendix, we show the residual variation for peers from the same cohort. So, to isolate peer effects, we use the fact that up to four immigrant teenagers from older cohorts do or do not take contraceptives compared to what could have been predicted by the use of fixed effects. This means that we narrow down the influence to very few (potentially close) peers.

Figure 2 here

The residual variation for native girls is plotted in figure 2. The order of magnitude of the residual variation is larger since much more native young women take contraceptives, both because numerically more natives attend high-school and because usage of contraceptives is more prevalent in this population. For girls from the older cohorts, there are more or less 0 to 50 young women who take contraceptives who could not have been predicted by relying on fixed effects. When assessing the influence of native schoolmates, we use a broader magnitude of change among peers.

5.3 Balancing tests

The identification strategy implies that characteristics that can explain the choice of school ($X_{i,c,s}$) should be uncorrelated to native and immigrant peers behaviour, conditional on the FE used to control for sorting. This motivates the following tests: regress each $x_{i,c,s}$ in $X_{i,c,s}$, meaning each of the variables used as controls, on peers behaviour, $\ddot{y}$ and the set of fixed effects used to control for sorting :

$$x_{i,c,s} = \alpha + \beta \ddot{y}_{c,s} + \lambda_s + \rho_s * c + u_{i,c,s} \tag{2.1}$$

and test $\mathbb{H}_0 : \beta = 0$. In equation 2.1, s stands for schools, c for cohorts and i for individuals, (λ_s) school fixed effects and $(\rho_s * c)$ school-specific time trends. Under no sorting, we should not reject $\mathbb{H}_0$. For instance, teenagers whose mother have a low level of education ($x_{i,c,s}$) can sort into schools with lower usage of contraceptives ($\ddot{y}$) generating a correlation between the two. If sorting is accounted for by using fixed effects, incorporating them should shrink this correlation to zero.

We run as many regressions as there are controls, definitions of influential groups and measures of peer behaviour. To measure behaviour, we want both to account for (i) the prevalence of

contraceptive usage in a group and (ii) the size of that group. We use three measures: the raw number of women taking contraceptives, the log of that number and the share of women in the population of reference conditional on group size. In the third case, we use size of the group as a separate regressor.

So, we run 17 regressions (number of controls) three times (number of measures), amounting to 51 specifications. We run them five times (number of peers behaviour), i.e. when we look at the usage of immigrant women in older cohorts, immigrant women from the same cohort, native women in older cohorts, native women in older cohorts and the composition of a cohort. In the last one, we look at the number of immigrants in the cohorts of first year students, its log and its share of all first year female students. This is because we also use size of the immigrant group as a regressor when prevalence of behaviour is measured as a share.

This amounts to 205 regressions. This has to be done separately for the peers of immigrants and of natives (since they do not go to exactly the same schools). This makes a total of 510 regressions. Since we run many regressions, some will necessarily reject $\mathbb{H}_0$. Table 7 reports how many times the null hypothesis is rejected at the 10, 5 and 1% level. The details of the regressions can be found in tables A3, A4 and A5.

Table 7 here

Overall, our rate of rejection is higher than it should be. We reject 14.9% of the time when we should reject 10%, 8.8% when we should reject 5% and 2.15% when we should regress 1%. This picture is not so different when looking at native and immigrant peers. There is over-rejection for all significance levels and the numbers of rejections are very close for the 10% level (39 for immigrants and 37 for natives).

A closer look at how the rejection rate breaks down by specifications gives a more optimistic picture. When focusing separately on the peers of natives and those of immigrants, there appears to be one specification which systematically over-rejects. For native peers, this is when we look at the influence of other native women from older cohorts. For immigrants, this is when we look at other immigrants from the same cohort. If we take these specifications out of the calculation, we return to the expected percentage of rejection for each significance level.

By considering table 7 without certain specifications, we do not cherry pick information. For each group of specification, we run 51 regressions. So, under the null, one should already expect rejection rates to correspond to the theoretical predictions for each subgroup. These results mean that we should be careful when interpreting the results of the influence of immigrant women from the same cohort on other immigrant teenagers and of older natives on younger ones.

6 Empirical Analysis

The strategy consists in regressing a dummy variable for having used contraceptives at least once by the end of the first, second or third year of high school ($y_{i,c,s}^{c+a}$) on a series of fixed effects that control for sorting ($\lambda_{c,s}$), individual characteristics ($X_{i,c,s}$) and peers behaviour ($\ddot{y}_{c,s}$). This estimated equation is:

$$y_{i,c,s}^{c+a} = \alpha + \beta \ddot{y}_{c,s} + \gamma X_{i,c,s} + \lambda_{c,s} + \epsilon_{i,c,s} \tag{2.2}$$

where subscript c refers to cohort, s to school, i to individuals and superscript a to an additional year of high school, meaning $a \in \{0, 1, 2\}$. The empirical choices boil down to (i) the choice of the set of fixed effects to control for sorting, (ii) the measure of peers behaviour (both how it is measured and which group to measure it for) and (iii) the variables which should serve as individual controls. It is estimated separately on immigrant and native teenagers.

Sorting is controlled for by including sequentially school fixed effects $\lambda_{s,c} = \omega_c$, school fixed effects and a time trend $\lambda_{s,c} = \omega_c + (c + a)$ and school fixed effects with a linear school specific trend $\lambda_{s,c} = \omega_c + \rho_s * (c + a)$. In the baseline analysis, influential peers are separately native and immigrant women from the second and third cohorts. Individual controls include the variables listed in table 5 and a dummy for being a first or a second generation immigrant. Equation 2.1 is estimated with linear least squares and standard errors are clustered at the school level.

Behaviour is measured as share of women in the group who has used contraceptives (together with a control for group size), raw number of women in the group who has used contraceptives and log of that number. The coefficients should be interpreted as follows, what variation (in percentage points) in the probability of using contraceptives would arise if: all women in the group were to take contraceptives (for a given group size), one more teenager with the characteristics of the group (native or immigrant) was to take contraceptives and what would happen if the number of teenagers with the characteristics of the group was to double (level - log regression).

6.1 Influence on non-Western immigrants

Regressions where the influential group is made of older immigrants and the influenced group is made of first year high-school teenagers with an immigrant background is shown in table 8. The first two columns report results from specifications where only year trends and year trends together with individual controls are included. Those are the naive estimators which do not account for sorting. It is hard to draw a picture from them since their sign is different depending on how peers behaviour is calculated. Thus, the direction of the bias is not straightforward.

Table 8 here

The causal effect (estimated in the other columns where we account for sorting), however, is

clearly negative, of large magnitude and long-lasting. The sign is negative for all measures of peers behaviour and sets of fixed effects used to account for sorting, in columns 3 to 6. Doubling the number of immigrants using contraceptives leads to a 5.1 percentage points decrease in the probability for an immigrant to take contraceptives in the first year (for a mean outcome of 13.63%), so a reduction of 37%. This is very large in magnitude, although lower than Card and Giuliano (2013) who found having friends starting to have sex leads to an almost 50% increase in the probability to become sexually active.

Under the most demanding specification, the coefficients are not significant, or only marginally, when peers behaviour is measured by the number of students using contraceptives. The effect is not precisely estimated because one extra immigrant taking contraceptives is a small change. Although the coefficients are stable after one, two or three years, the mean outcome increases between the beginning and the end of high-school. The effect in percentage points is similar but it represents in proportion less and less with time showing that the effect dies out. We take these results as evidence that immigrant women are influenced by other immigrant peers and that this influence is negative.

In table 9, we reproduce the empirical analysis to study the influence of older native students on younger immigrant ones. The picture is very different from the one that emerged in the previous table. The naive estimators in column 1 and 2 are positive while the causal estimates are zero. When measured with fraction of usage and number of teenagers having used contraceptives, the estimates are small in magnitude and insignificant. The sign switches between specification when the regressor is a share showing that the estimates are not different from zero. Overall, they do point towards a negative effect and are even marginally significant after two years for the log measure but they are very small.

One cannot compare the effects of doubling the number of natives and immigrants taking contraceptives. The two groups have too different sizes. However, if all natives were to take contraceptives, immigrants would not respond (panel A of table 9), while they would if all immigrants suddenly took contraceptives. We take it as evidence that young immigrants are not influenced by their native schoolmates.

Table 9 here

A potential criticism to our approach is that so far we pool together all non-Western origins as if it was a homogeneous group. We should make a difference between immigrant groups based on how culturally close they are. It is however too restrictive to focus on peers from the same country of origin as there are too few in our sample and residual variation is likely to be very small. We thus create a dummy for coming from a mostly Muslim country and reproduce the analysis to study Muslim to Muslim (table 10) and non-Muslim to Muslim (table 11) influence. Table 10 confirms the results of table 8. The coefficients are negative and compared to table 8 are even larger in terms of size, making it statistically significant for all specifications for year 1 and

2 and almost all for year 3. An additional Muslim girl taking the pill is associated with a 1.82 percentage points decrease in the probability for younger girls to take contraceptives in the first year. Similarly to table 9, there does not seem to be any influence from older non-Muslim students on younger Muslim ones.

Tables 10 and 11 here

6.2 Influence on native teenagers

We reproduce the analysis on peer influence for natives. This allows us to see if the negative result on immigrants, i.e. a lowering of the probability to take contraceptives, is atypical or if it also holds for the majority group. The results on natives to natives influence are reported in table 12. They also clearly show a negative effect. If all native girls in older cohorts were to take contraceptives, this would reduce the probability for a first year native of 7.11 percentage points, which represent a 12% decrease. The magnitude is lower than for immigrants but the two groups have very different mean usage in the first place so it is not expected that the effect is quantitatively the same. They show that the negative effect is consistent across groups.

Table 12 here

We also enquire whether there is influence of older immigrant women on natives. Results of this analysis are reported in table 13. While immigrant peers are less influential than their native counterparts, younger Danes without foreign background are not closed to the influence of immigrants. The effect is negative and significant for the first year of high school. This asymmetric result compared to how immigrants see natives is not driven by sample size since the coefficients from table 13 are larger than those from tables 9 and 11. We take it as evidence that although less clear cut, there is influence of immigrant peers on natives. There is therefore a difference between how immigrants treat the information from natives and how natives treat the information form immigrants.

Table 13 here

7 Evidence on Mechanisms

Once this negative result on usage of contraceptives is established, it is important to dig into explaining it. In particular, we are interested to see if immigrant change their choice of contraception only because they change their medication or because they change their sexual behaviour.

To do so, we reproduce the baseline analysis but instead of using contraceptive usage as an outcome, we look at probability of having an abortion, probability of being treated for chlamydia

and to have had a consultation with a GP. If medication alone is affected by peers behaviour, we expect to see a negative effect on the probability of having a consultation but no effect on abortion and chlamydia (and potentially a positive effect on abortion). If sexual behaviours change, one should expect a decrease in the probability of having an abortion and/or of being treated for chlamydia.

Results on the influence of older immigrants on younger ones are reported in table 14. The table is divided in three groups of columns, each showing the results for one of the school year. To not clutter the tables, we do not report the naive estimates that do not account for sorting.

Table 14 here

There is a clear negative effect on the probability of having a consultation with a GP. A doubling of the number of immigrant taking the pill leads to a 3.39 percentage points decrease in the probability to have a consultation in the first year. This result is not surprising as this outcome is highly correlated with usage of contraceptives. It acts nevertheless as an indirect robustness check.

We also interpret the effect on abortion to be negative. Although most coefficients are not statistically significant, all specifications and all measures of peers behaviour produce a negative coefficient. The coefficients themselves are relatively large since doubling the number of older immigrant taking contraceptives causes a 0.31 percentage points decrease in the probability of having an abortion in the first year of high-school, for a mean outcome of 0.94% (among immigrants in the first year of high school).

We interpret the effect on chlamydia to be null. All specifications but one are not statistically significant but most importantly the sign of the coefficients change with the specification and the measure of peers behaviour used, both between and within years. Our conclusion is that peers behaviour lower the probability to take contraceptives partially by changing immigrant sexual behaviour, at least measured by the probability of having an abortion and not only through the effect on medication.

Table 15 here

We then look at the effects of native peers on younger immigrants to see if the absence of influence by native schoolmates is confirmed when one looks at other outcome variables. Results are reported in table 15 and point to the absence of effect of native peers. Almost all coefficients are statistically insignificant but most importantly they change sign between measures and specifications and are very small in magnitude. This last concern is especially relevant when peers behaviour is measured with the number of women using contraceptives. For the outcomes consultation and abortion, the sign of the coefficient changes between the log and the share measures which we take as indicative of a null effect. For chlamydia, the sign of the coefficient seems positive

when measured with logs but change sign between specifications when measured with a fraction. We also interpret it as evidence of no effect. This reinforces the conclusion on the absence of influence of native peers.

To make sure that the patterns we observe are not exclusive to immigrants, we reproduce this analysis focusing this time on natives to natives influence. Results are reported in table 16. The results appear negative for the outcome consultation confirming the picture depicted in table 12 and that peer effects among immigrants are not atypical. If all older natives were to take contraceptives, the probability of having a consultation for a younger native would decrease by 4.13 percentage points (for a mean outcome àf 56.39%). The result for abortion point to a negative effect more pronounced in the second and third years of high school. The results for chlamydia appear highly significant statistically when peers behaviour is measured by the number of women taking contraceptives. However, they change signs between specifications, and also when measured with logs. The sample size is very large and very small differences can be statistically different from zero. It is thus not possible to detect a consistent sign for the outcome chlamydia.

Table 16 here

Our take-away is that peers influence go, at least partially, through a change in sexual behaviour (in particular a decrease in the probability of having an abortion). This channel is not atypical of immigrants but goes only through immigrant influence on other immigrants.

8 Robustness Checks

The main robustness check we conduct, is to see if the results change when instead of looking at older peers, we look at the behaviour of peers from the same cohort. We isolate a residual variation in the (pre high-school) behaviour of first year students and see how it influences their schoolmates in the first, second and third years of high school.

Results are reported in table 17 for the influence of immigrants and in table 18 of natives on immigrants. They confirm the picture established when using older girls as potential role models. Influence from other immigrants cause a decrease in the use of contraceptives while changes in the behaviour of natives have no effect.

Tables 17 and 18 here

9 Conclusion

In this paper, we look at teenage peers influence over the decision to take contraceptives. Since using contraception is almost universal in the population of teenage natives in Western Europe but not among young immigrants with a non-Western background, we use it as a measure of cultural

adaptation. We exploit residual variation in usage of contraceptives by older schoolmates to identify peer effects. Looking at the separate influence of native and immigrant peers is indicative of who has influence in immigrants process of cultural adaptation.

The more older immigrant peers take contraceptives, the less younger ones use them. This negative effect is not atypical since younger natives also respond negatively to an increase in contraceptive usage by older native schoolmates. Immigrants are however closed to the influence of natives, in the sense that they do not change their behaviour in response to what older natives do. This asymmetric influence does not hold in the population of young natives who are influenced by members of their own groups and also by immigrants.

Bibliography

Altonji, J. G., Cattan, S., and Ware, I. (2017). Identifying Sibling Influence on Teenage Substance Use. *Journal of Human Ressources*, 1(52):1–47.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30(C):98–108.

Angrist, J. D. and Lang, K. (2004). Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *American Economic Review*, 94(5):1613–1634.

Argys, L. M. and Rees, D. (2008). Searching for peer group effects: A test of the contagion hypothesis. *The Review of Economics and Statistics*, 90(3):442–458.

Bailey, M. J. (2006). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply. *The Quarterly Journal of Economics*, 121(1):289–320.

Ballatore, R. M., Fort, M., and Ichino, A. (2018). Tower of Babel in the Classromm: Immigrants and Natives in Italian Schools. *Journal of Labor Economics*, 36(4):885–921.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2008). Are Muslim Immigrants Different in Terms of Cultural Integration? *Journal of the European Economic Association*, 6(2-3):445–456.

Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like Beckham: Ethnic identity and integration. *European Economic Review*, 90(C):146–164.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, 31(1):119–153.

Brenoe, A. and Zolitz, U. (2018). Exposure to more female peers widens the gender gap in stem participation. Technical Report 285, Zurich Working Paper Series.

Card, D. and Giuliano, L. (2013). Peer Effects and Multiple Equilibria in the Risky Behavior of Friends. *The Review of Economics and Statistics*, 95(4):1130–1149.

Clark, A. E. and Loheac, Y. (2007). It wasn t me, it was them! social influence in risky behavior by adolescents. *Journal of Health Economics*, 26(4):763 – 784.

Goldin, C. and Katz, L. F. (2002). The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions. *Journal of Political Economy*, 110(4):730–770.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. NBER Working Papers 7867, National Bureau of Economic Research, Inc.

Kuran, T. and Sandholm, W. (2008). Cultural Integration and Its Discontents. *Review of Economic Studies*, 75(1):201–228.

Lalonde, R. and Giguere, B. (2008). When might the two cultural worlds of second generation biculturals collide? *Canadian Diversity*, 6(2):58–62.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3):531–542.

Merlino, L., Steinhardt, M., and Wren-Lewis, L. More than just friends? school peers and adult interracial relationships. *Journal of Labor Economics*. forthcoming.

Olcina, G., Panebianco, F., and Zenou, Y. (2017). Conformism, Social Norms and the Dynamics of Assimilation. CEPR Discussion Papers 12166, C.E.P.R. Discussion Papers.

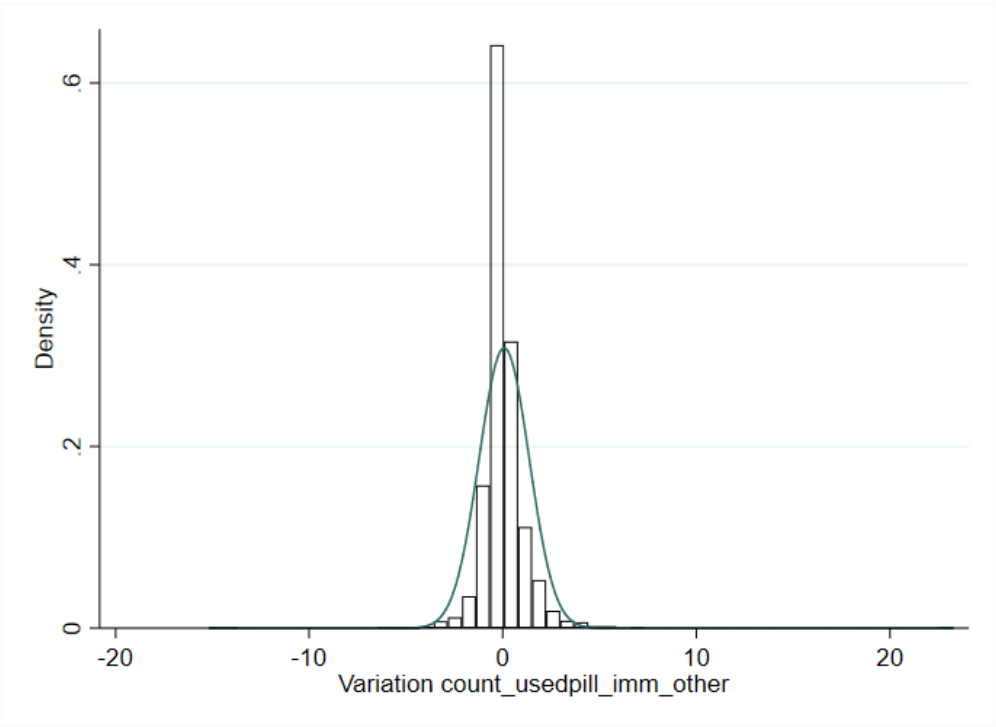
Olivetti, C., Patacchini, E., and Zenou, Y. Mothers, peers and gender identity. *Journal of the European Economic Association*. forthcoming.

Patacchini, E. and Zenou, Y. (2016). Social networks and parental behavior in the intergenerational transmission of religion. *Quantitative Economics*, 7(3):969–995.

Verdier, T. and Zenou, Y. (2017). The role of social networks in cultural assimilation. *Journal of Urban Economics*, 97(C):15–39.

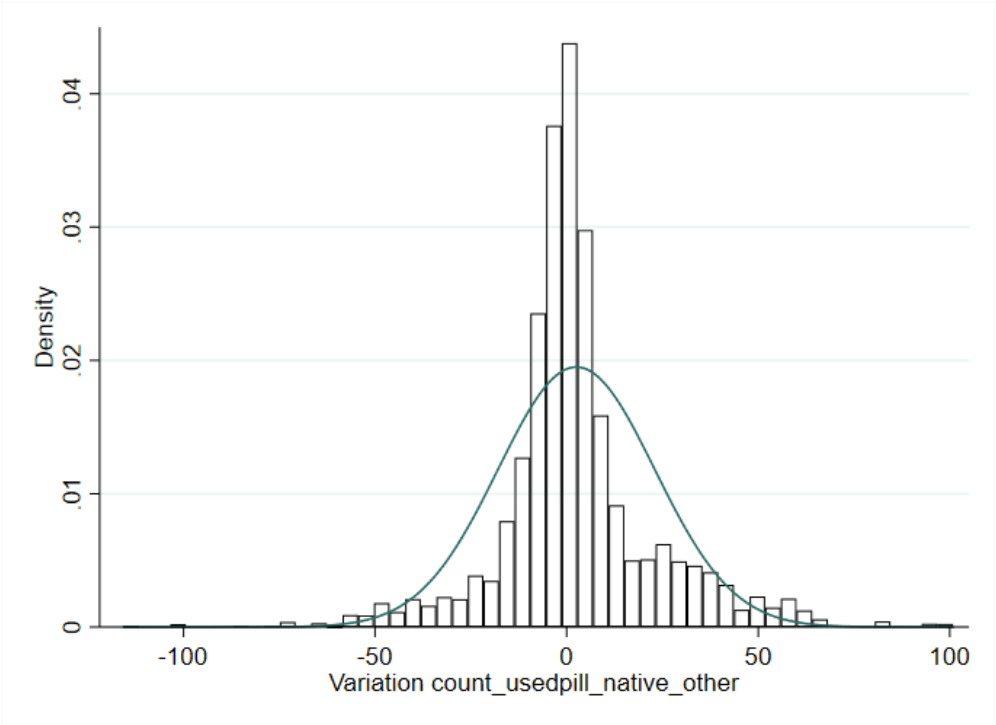
Verdier, T. and Zenou, Y. (2018). Cultural leader and the dynamics of assimilation. *Journal of Economic Theory*, 175(C):374–414.

Figure 1: Residual variation for identification (immigrant women from older cohorts)



Note : This graph plots the residuals of a regression of the number of immigrant women from older cohorts (by school/cohort) who have taken contraceptives on school fixed effects and a school specific time trend. The residual is the difference between the predicted and actual numbers (at the school/cohort level). The curve is a fitted normal distribution.

Figure 2: Residual variation for identification (native women from older cohorts)



Note : This graph plots the residuals of a regression of the number of native women from older cohorts (by school/cohort) who have taken contraceptives on school fixed effects and a school specific time trend. The residual is the difference between the predicted and actual numbers (at the school/cohort level). The curve is a fitted normal distribution.

Table 1: Distribution of schools - program and student level

	Program Level	School level			
	% schools	% schools	% non Western (mean)	% non Western (std)	Nb students
STX - General (Math)	21.26	19.24	5.72	7.75	41,043
STX - General (Language)	21.11	21.70	4.53	5.94	46,290
STX - General (Combined)	22.50	38.71	7.17	8.03	82,576
HF - 2 years General	5.24	0.66	5.55	6.37	1,405
HTX - Business	15.10	3.52	6.79	7.01	7,512
HHX - Technical	14.79	16.18	5.61	5.34	34,520
Overall			6.04	7.20	213,346
Nb observations	649	213,346			

Note: This table reports the distribution of programs and students among schools, together with information on the share of non-Western immigrants. The first column reports the distribution of programs at the school level. The second column reports the share of students going to each type of school, while the third and fourth columns report the mean and standard deviation of the share of non-Western immigrants. The last column reports the number of students going to this program. The second row should be read as follows, 21.11% of the 643 schools are General (STX) Language schools. 46,290 students, so 21.70% of the entire sample attend these schools, 4.53% of them are non-Western immigrants (first and second generation), for a standard deviation of 0.0775

Table 2: Distribution of Immigrants - By Origin country and Religious Affiliation

	Number of obs	% of sample
Panel A: Western/non-Western background		
Western Background	201,264	94.34
non-Western immigrants	12,082	5.66
Panel B: Religious Affiliation		
non-Muslim	205,628	96.38
Muslim	7,718	3.62
Panel C: Country of origin - non-Western		
Yugoslavia	541	4.48
Turkey	2,368	19.60
Morocco	379	3.14
Somalia	275	2.28
Afghanistan	572	4.73
Sri Lanka	668	5.53
Iraq	839	6.94
Iran	695	5.75
China	167	1.38
Lebanon	943	7.80
Pakistan	1,068	8.84
Vietnam	601	4.97
Russia	151	1.25
Bosnia-Herzegovina	1,073	8.88
Panel C: Country of origin - Western		
Denmark	199,323	99.04
Iceland	162	0.08
Netherlands	161	0.08
Poland	576	0.29
Germany	251	0.12

Note: This table reports the distribution of students by origin, both the number of observations and the share of the sample. Panel A looks at the distinction Western/non-Western background. Panel B breaks down the population between originating from a mostly Muslim country or not. Panel C breaks down country of origin for students with a non-Western background and panel D does the same for students with a Western background. This table only reports countries from which more than 150 students are observed.

Table 3: Outcomes - Difference Natives and Immigrants by age

Age	15	16	17	18	19	20
Contraception, natives	0.2484	0.4554	0.6266	0.7424	0.8116	0.8573
	(0.43)	(0.50)	(0.48)	(0.44)	(0.39)	(0.35)
Contraception, immigrants	0.0478	0.1009	0.1669	0.2506	0.3335	0.4081
	(0.21)	(0.30)	(0.37)	(0.43)	(0.47)	(0.49)
Abort, natives	0.0038	0.0089	0.0160	0.0255	0.0373	0.0491
	(0.06)	(0.09)	(0.13)	(0.16)	(0.19)	(0.22)
Abort, immigrants	0.0020	0.0049	0.0102	0.0226	0.0378	0.0545
	(0.04)	(0.07)	(0.10)	(0.15)	(0.19)	(0.23)
Chlamydia, natives	0.0115	0.0247	0.0484	0.0813	0.1168	0.1511
	(0.11)	(0.16)	(0.21)	(0.27)	(0.32)	(0.36)
Chlamydia, immigrants	0.0045	0.0078	0.0137	0.0250	0.0354	0.0489
	(0.07)	(0.09)	(0.12)	(0.16)	(0.18)	(0.22)
Consultation, natives	0.2362	0.4251	0.5777	0.6792	0.7418	0.7868
	(0.42)	(0.49)	(0.49)	(0.47)	(0.44)	(0.41)
Consultation, immigrants	0.0573	0.1138	0.1805	0.2602	0.3398	0.4083
	(0.23)	(0.32)	(0.38)	(0.44)	(0.47)	(0.49)
Gynaecologist, natives	0.0127	0.0251	0.0431	0.0655	0.0924	0.1232
	(0.11)	(0.16)	(0.20)	(0.25)	(0.29)	(0.33)
Gynaecologist, immigrants	0.0161	0.0252	0.0415	0.0573	0.0821	0.1128
	(0.13)	(0.16)	(0.20)	(0.23)	(0.27)	(0.32)

Note: This table reports information on the sexual behaviour of teenagers and young women (who went to high school) in Denmark. It details the share who have used contraception at least once for all ages between 15 and 20 years old, together with the share who ever had an abortion, were ever treated for chlamydia, ever saw a gynaecologist or had a consultation with their GP to specifically talk about contraception. The category natives relates to young women without foreign background, while the category immigrants refers to non-Western immigrants (first and second generation).

Table 4: Outcomes - Difference Natives and Immigrants by school level

Years in high school	Natives		Immigrants	
Contraception, year 1	0.5639	(0.50)	0.1363	(0.34)
Contraception, year 2	0.6993	(0.46)	0.2105	(0.41)
Contraception, year 3	0.7834	(0.41)	0.2926	(0.45)
Abort, year 1	0.0137	(0.12)	0.0094	(0.10)
Abort, year 2	0.0222	(0.15)	0.0187	(0.14)
Abort, year 3	0.0332	(0.18)	0.0327	(0.18)
Chlamydia, year 1	0.0394	(0.19)	0.0116	(0.11)
Chlamydia, year 2	0.0684	(0.25)	0.0188	(0.14)
Chlamydia, year 3	0.1030	(0.30)	0.0305	(0.17)
Consultation, year 1	0.5250	(0.50)	0.1529	(0.36)
Consultation, year 2	0.6419	(0.48)	0.2245	(0.42)
Consultation, year 3	0.7147	(0.45)	0.2980	(0.46)
Gynaecologist, year 1	0.0348	(0.18)	0.0329	(0.18)
Gynaecologist, year 2	0.0549	(0.23)	0.0493	(0.22)
Gynaecologist, year 3	0.0796	(0.27)	0.0692	(0.25)

Note: This table reports information on the sexual behaviour of teenagers and young women in Denmark. It details the share who have used contraception at least once by the beginning of the three different years of high school. It also shows the share who ever had an abortion, were ever treated for chlamydia, ever saw a gynaecologist or had a consultation with their GP to specifically talk about contraception. The category natives relates to young women without foreign background, while the category immigrants refers to non-Western immigrants (first and second generation).

Table 5: Background variables - Individual level

	Natives		Immigrants	
Parents married or cohabiting	0.74	(0.44)	0.79	(0.40)
Unemployed, mom	0.03	(0.16)	0.11	(0.31)
Out of the labour force, mom	0.08	(0.27)	0.48	(0.50)
Missing employment status, mom	0.01	(0.09)	0.02	(0.14)
Months of education, mom	153.45	(36.54)	66.34	(72.79)
Education above high school, mom	0.40	(0.49)	0.10	(0.30)
Missing education, mom	0.02	(0.15)	0.50	(0.50)
Age at birth, mom	28.39	(4.60)	25.38	(6.06)
Missing birth information, mom	0.00	(0.02)	0.01	(0.11)
Unemployed, dad	0.02	(0.14)	0.09	(0.29)
Out of the labour force, dad	0.04	(0.20)	0.31	(0.46)
Missing employment status, dad	0.04	(0.19)	0.11	(0.32)
Months of education, dad	150.85	(47.24)	72.78	(77.11)
Education above high school, dad	0.33	(0.47)	0.14	(0.34)
Missing education, dad	0.06	(0.23)	0.50	(0.50)
Age at birth, dad	30.76	(6.36)	27.84	(10.67)
Missing birth information, dad	0.01	(0.11)	0.08	(0.28)
Age at high school start	16.14	(0.73)	16.09	(0.85)
Nb observations	201,264		14,284	

Note: This table reports average characteristics together with their standard deviation for young women both without a foreign background (natives) and who are first or second generation immigrants. Characteristics are expressed at the individual level, meaning the row "Age at birth, mom" should be read as follows, the average age at birth of mothers of the 201,264 natives is 28.39 years old when it is 25.28 years old for the 14,284 immigrants.

Table 6: Background variables - Cohort characteristics

	Natives		Immigrants	
Fraction employed mother	0.84	(0.07)	0.79	(0.11)
Fraction employed father	0.82	(0.07)	0.77	(0.10)
Avg. Months of educ. Mother	148.01	(11.44)	139.33	(17.06)
Fraction above high school mother	0.41	(0.14)	0.38	(0.13)
Avg. Months of educ. Father	137.94	(11.81)	129.14	(16.85)
Fraction above high school father	0.33	(0.13)	0.31	(0.11)
Average age at birth, mom	28.05	(1.10)	27.91	(1.06)
Average age at birth, dad	28.68	(1.32)	28.44	(1.30)
Fraction, parents married or cohabiting	0.70	(0.07)	0.68	(0.07)
Fraction missing employment status, mother	0.03	(0.04)	0.03	(0.06)
Fraction missing employment status, father	0.11	(0.05)	0.13	(0.07)
Fraction missing education status, mother	0.06	(0.05)	0.11	(0.09)
Fraction missing education status, father	0.14	(0.06)	0.19	(0.09)
Fraction missing information, mom	0.01	(0.01)	0.01	(0.01)
Fraction missing information, dad	0.08	(0.04)	0.09	(0.04)
Fraction of boys immigrants	0.06	(0.07)	0.14	(0.13)
Fraction of girls immigrants	0.05	(0.07)	0.15	(0.15)
Average students	155.64	(94.15)	169.09	(102.89)
Frac used the pill in grade of girls, natives	0.36	(0.14)	0.37	(0.13)
Frac used the pill in grade of girls, immigrants	0.08	(0.18)	0.07	(0.14)
Frac used the pill older girls, natives	0.57	(0.18)	0.56	(0.18)
Frac used the pill older girls, immigrants	0.16	(0.23)	0.14	(0.17)

Note: This table reports average characteristics together with their standard deviation for the cohorts where young women, both without a foreign background (natives) and who are first or second generation immigrants, study. The row average students should be reads as follows: in cohorts where there are native students there are on average 155.64 students (both men and women) in a cohort, while in cohorts where there are immigrant students, there are on average 169.09 students.

Table 7: Summary of balancing tests

Category	Number of regressions	Number of rejections		
		10% level	5% level	1% level
Cohort Composition - Immigrants	51	9	2	0
Immigrant women in same grade - Immigrants	51	15	11	3
Immigrant women in older grades - Immigrants	51	3	2	0
Native women in same grade - Immigrants	51	5	2	0
Native women in older grades - Immigrants	51	7	2	0
Cohort Composition - Natives	51	6	5	1
Immigrant women in same grade - Natives	51	3	2	0
Immigrant women in older grades - Natives	51	7	4	0
Native women in same grade - Natives	51	7	5	1
Native women in older grades - Natives	51	14	10	6
Cohort Composition	102	15	17	1
Immigrant women in same grade	102	18	13	3
Immigrant women in older grades	102	10	6	0
Native women in same grade	102	12	7	1
Native women in older grades	102	21	12	6
Fraction	170	22	11	1
Log	170	27	16	5
Numbers	170	27	18	5
Immigrants	255	39	19	3
Natives	255	37	26	8
Total	510	76	45	11

Note: This table provide a summary of the balancing tests. For each grouping of regressions, i.e. each row, the second column reports the number of regressions performed and in the following columns the number of rejection of $\mathbb{H}_0$ at the 10, 5 and 1% significance level.

Table 8: Influence of older immigrant women on immigrant teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.105***	0.0918***	-0.153***	-0.157***	-0.153***	-0.283***
	(0.0238)	(0.0227)	(0.0267)	(0.0267)	(0.0261)	(0.0296)
Year 2	0.129***	0.109***	-0.173***	-0.177***	-0.174***	-0.302***
	(0.0273)	(0.0262)	(0.0312)	(0.0313)	(0.0307)	(0.0345)
Year 3	0.165***	0.141***	-0.141***	-0.145***	-0.142***	-0.250***
	(0.0302)	(0.0289)	(0.0348)	(0.0346)	(0.0339)	(0.0397)
			Panel B: Numbers of women using contraceptives			
Year 1	-0.00167	-0.00117	-0.00278**	-0.00406***	-0.00366**	-0.00653
	(0.00112)	(0.000823)	(0.00115)	(0.00154)	(0.00157)	(0.00399)
Year 2	-0.00252**	-0.00170*	-0.00367***	-0.00507***	-0.00459***	-0.00800*
	(0.00124)	(0.000864)	(0.00139)	(0.00178)	(0.00176)	(0.00446)
Year 3	-0.00286	-0.00171	-0.00271	-0.00401*	-0.00350*	-0.00671
	(0.00196)	(0.00134)	(0.00169)	(0.00205)	(0.00198)	(0.00436)
			Panel C: Log of numbers			
Year 1	-0.0120**	-0.00745	-0.0147**	-0.0231***	-0.0210***	-0.0509***
	(0.00572)	(0.00521)	(0.00595)	(0.00608)	(0.00596)	(0.00765)
Year 2	-0.0170**	-0.00997	-0.0209***	-0.0307***	-0.0280***	-0.0619***
	(0.00670)	(0.00615)	(0.00726)	(0.00760)	(0.00735)	(0.00976)
Year 3	-0.0228***	-0.0135**	-0.0175**	-0.0267***	-0.0237***	-0.0553***
	(0.00799)	(0.00680)	(0.00803)	(0.00846)	(0.00812)	(0.0107)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools			583	583	583	583
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of non-Western women from the second and the third years of high school on non-Western immigrants from the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 9: Influence of older native women on immigrant teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.0628***	0.0633***	0.0112	0.000310	-0.00568	-0.0177
	(0.0216)	(0.0205)	(0.0250)	(0.0260)	(0.0251)	(0.0333)
Year 2	0.0850***	0.0856***	0.0232	0.00948	0.00164	-0.00920
	(0.0256)	(0.0239)	(0.0317)	(0.0332)	(0.0312)	(0.0436)
Year 3	0.0987***	0.101***	0.0359	0.0239	0.0168	0.0388
	(0.0275)	(0.0255)	(0.0328)	(0.0343)	(0.0322)	(0.0494)
			Panel B: Numbers of women using contraceptives			
Year 1	0.000434***	0.000396***	6.33e-05	-3.06e-05	2.21e-05	-0.000127
	(8.53e-05)	(8.10e-05)	(9.71e-05)	(0.000107)	(0.000107)	(0.000174)
Year 2	0.000545***	0.000488***	-2.24e-05	-0.000148	-7.57e-05	-0.000223
	(0.000106)	(9.85e-05)	(0.000122)	(0.000138)	(0.000136)	(0.000234)
Year 3	0.000615***	0.000549***	2.30e-05	-8.81e-05	-3.26e-06	-9.32e-05
	(0.000106)	(9.26e-05)	(0.000119)	(0.000137)	(0.000135)	(0.000210)
			Panel C: Log of numbers			
Year 1	0.0101***	0.00882***	-4.72e-05	-0.00332	-0.00212	-0.00499
	(0.00317)	(0.00298)	(0.00292)	(0.00311)	(0.00311)	(0.00402)
Year 2	0.0120***	0.00994***	-0.00383	-0.00822**	-0.00659*	-0.00841*
	(0.00393)	(0.00358)	(0.00369)	(0.00404)	(0.00394)	(0.00507)
Year 3	0.0175***	0.0152***	-0.00183	-0.00569	-0.00370	-0.00310
	(0.00421)	(0.00375)	(0.00390)	(0.00432)	(0.00417)	(0.00540)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools			583	583	583	583
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of native women from the second and the third years of high school on non-Western immigrants from the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 10: Influence of older Muslim women on Muslim teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.0333	0.0262	-0.155***	-0.160***	-0.161***	-0.328***
	(0.0247)	(0.0241)	(0.0317)	(0.0320)	(0.0319)	(0.0398)
Year 2	0.0565*	0.0435	-0.198***	-0.202***	-0.206***	-0.367***
	(0.0304)	(0.0301)	(0.0363)	(0.0367)	(0.0366)	(0.0452)
Year 3	0.0915**	0.0769**	-0.192***	-0.195***	-0.199***	-0.329***
	(0.0359)	(0.0358)	(0.0412)	(0.0410)	(0.0413)	(0.0569)
			Panel B: Numbers of women using contraceptives			
Year 1	-0.00493***	-0.00415***	-0.00457**	-0.00729***	-0.00738***	-0.0182***
	(0.00128)	(0.00134)	(0.00190)	(0.00199)	(0.00201)	(0.00360)
Year 2	-0.00561***	-0.00445**	-0.00508**	-0.00728***	-0.00749***	-0.0187***
	(0.00191)	(0.00196)	(0.00235)	(0.00255)	(0.00246)	(0.00454)
Year 3	-0.00528*	-0.00375	-0.00311	-0.00492	-0.00518	-0.0152***
	(0.00272)	(0.00270)	(0.00333)	(0.00349)	(0.00329)	(0.00478)
			Panel C: Log of numbers			
Year 1	-0.0145***	-0.0119**	-0.00968	-0.0174**	-0.0176**	-0.0446***
	(0.00443)	(0.00464)	(0.00670)	(0.00696)	(0.00703)	(0.0105)
Year 2	-0.0173***	-0.0133**	-0.0126	-0.0188**	-0.0194**	-0.0453***
	(0.00630)	(0.00648)	(0.00783)	(0.00851)	(0.00835)	(0.0130)
Year 3	-0.0173**	-0.0122	-0.00737	-0.0126	-0.0134	-0.0373***
	(0.00878)	(0.00878)	(0.0107)	(0.0112)	(0.0107)	(0.0135)
Nb of observations	7,718	7,718	7,718	7,718	7,718	7,718
Nb of schools			516	516	516	516
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of women from a Muslim background in the second and the third years of high school on immigrants with a Muslim background in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 11: Influence of older non-Muslim women on Muslim teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.0390*	0.0425**	0.0402	0.0299	0.0254	0.0154
	(0.0213)	(0.0212)	(0.0254)	(0.0264)	(0.0260)	(0.0348)
Year 2	0.0798***	0.0838***	0.0470	0.0408	0.0344	0.0227
	(0.0249)	(0.0247)	(0.0295)	(0.0304)	(0.0298)	(0.0409)
Year 3	0.107***	0.113***	0.0539	0.0507	0.0440	0.0714
	(0.0278)	(0.0277)	(0.0344)	(0.0356)	(0.0351)	(0.0494)
			Panel B: Numbers of women using contraceptives			
Year 1	0.000183***	0.000165***	-1.42e-05	-0.000145	-0.000140	-0.000177
	(6.25e-05)	(6.15e-05)	(0.000105)	(0.000118)	(0.000119)	(0.000205)
Year 2	0.000327***	0.000301***	-2.52e-05	-0.000120	-0.000111	-0.000138
	(8.63e-05)	(8.24e-05)	(0.000114)	(0.000129)	(0.000129)	(0.000225)
Year 3	0.000452***	0.000420***	7.38e-05	1.69e-05	2.47e-05	7.29e-06
	(9.09e-05)	(8.59e-05)	(0.000124)	(0.000140)	(0.000141)	(0.000255)
			Panel C: Log of numbers			
Year 1	0.00339	0.00246	-0.000581	-0.00420	-0.00448	-0.00270
	(0.00296)	(0.00295)	(0.00312)	(0.00344)	(0.00347)	(0.00455)
Year 2	0.00850**	0.00706**	-0.000829	-0.00342	-0.00380	-0.00194
	(0.00342)	(0.00335)	(0.00357)	(0.00398)	(0.00396)	(0.00515)
Year 3	0.0148***	0.0131***	0.00204	0.000346	-0.000139	0.00153
	(0.00384)	(0.00374)	(0.00396)	(0.00435)	(0.00433)	(0.00580)
Nb of observations	7,718	7,718	7,718	7,718	7,718	7,718
Nb of schools			516	516	516	516
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of women from a non-Muslim background in the second and the third years of high school on immigrants with a Muslim background in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 12: Influence of older native women on native teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.390***	0.304***	0.0816***	-0.00654	-0.00542	-0.0711***
	(0.0226)	(0.0183)	(0.0136)	(0.0113)	(0.0111)	(0.0144)
Year 2	0.317***	0.247***	0.0413***	-0.0188*	-0.0178	-0.0799***
	(0.0196)	(0.0165)	(0.0119)	(0.0110)	(0.0108)	(0.0133)
Year 3	0.241***	0.189***	0.0155	-0.0205**	-0.0198**	-0.0668***
	(0.0162)	(0.0142)	(0.00967)	(0.00931)	(0.00922)	(0.0116)
			Panel B: Numbers of women using contraceptives			
Year 1	0.000541***	0.000430***	0.000269***	-0.000233***	-0.000250***	-0.000443***
	(9.27e-05)	(7.40e-05)	(3.61e-05)	(4.02e-05)	(3.89e-05)	(7.01e-05)
Year 2	0.000435***	0.000342***	0.000132***	-0.000208***	-0.000219***	-0.000313***
	(7.20e-05)	(5.71e-05)	(3.10e-05)	(3.75e-05)	(3.70e-05)	(6.41e-05)
Year 3	0.000362***	0.000294***	8.00e-05***	-0.000121***	-0.000128***	-0.000188***
	(5.32e-05)	(4.32e-05)	(2.64e-05)	(3.10e-05)	(3.04e-05)	(5.24e-05)
			Panel C: Log of numbers			
Year 1	0.0173***	0.0140***	0.00937***	-0.00495***	-0.00546***	-0.00437***
	(0.00300)	(0.00243)	(0.00119)	(0.00130)	(0.00125)	(0.00168)
Year 2	0.0142***	0.0114***	0.00483***	-0.00477***	-0.00510***	-0.00327**
	(0.00252)	(0.00209)	(0.00107)	(0.00116)	(0.00113)	(0.00157)
Year 3	0.0122***	0.0101***	0.00273***	-0.00298***	-0.00318***	-0.00158
	(0.00205)	(0.00176)	(0.000864)	(0.000922)	(0.000902)	(0.00128)
Nb of observations	201,264	201,264	201,264	201,264	201,264	201,264
Nb of schools			647	647	647	647
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of native women in the second and the third years of high school on native teenagers background in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 13: Influence of older immigrant women on native teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.0608***	0.0493***	4.44e-05	-0.0131**	-0.0134**	-0.0151**
	(0.0125)	(0.00992)	(0.00717)	(0.00625)	(0.00615)	(0.00713)
Year 2	0.0458***	0.0364***	-0.00518	-0.0138**	-0.0140**	-0.0126*
	(0.0102)	(0.00834)	(0.00665)	(0.00608)	(0.00602)	(0.00665)
Year 3	0.0371***	0.0302***	-0.00256	-0.00769	-0.00789	-0.00553
	(0.00810)	(0.00681)	(0.00551)	(0.00524)	(0.00518)	(0.00602)
			Panel B: Numbers of women using contraceptives			
Year 1	0.00490***	0.00348**	0.00516***	-0.00188**	-0.00206**	-0.00180*
	(0.00185)	(0.00160)	(0.00157)	(0.000797)	(0.000815)	(0.00104)
Year 2	0.00351**	0.00238*	0.00296***	-0.00167**	-0.00181**	-0.000778
	(0.00147)	(0.00127)	(0.00109)	(0.000718)	(0.000732)	(0.000823)
Year 3	0.00275**	0.00192*	0.00181*	-0.000923	-0.00101	-0.000421
	(0.00118)	(0.00104)	(0.00106)	(0.000676)	(0.000654)	(0.000828)
			Panel C: Log of numbers			
Year 1	0.0201**	0.0142**	0.0210***	-0.00644*	-0.00745**	-0.00705*
	(0.00823)	(0.00686)	(0.00337)	(0.00340)	(0.00334)	(0.00407)
Year 2	0.0150**	0.0104*	0.0132***	-0.00478	-0.00553*	-0.00151
	(0.00651)	(0.00540)	(0.00295)	(0.00309)	(0.00305)	(0.00387)
Year 3	0.0117**	0.00830*	0.00875***	-0.00178	-0.00229	0.000434
	(0.00499)	(0.00423)	(0.00258)	(0.00267)	(0.00263)	(0.00352)
Nb of observations	201,264	201,264	201,264	201,264	201,264	201,264
Nb of schools			647	647	647	647
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of non-Western immigrant women in the second and the third years of high school on native teenagers in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

Table 14: Influence of older immigrant women on immigrant teenagers - Mechanisms

	Year 1				Year 2				Year 3			
					Panel A: Fraction of usage							
Abortion	-0.0110	-0.0113	-0.0112	-0.0183	-0.00549	-0.00602	-0.00562	-0.0153	-0.000807	-0.00203	-0.00153	-0.0211
	(0.00882)	(0.00880)	(0.00883)	(0.0116)	(0.0115)	(0.0116)	(0.0116)	(0.0156)	(0.0152)	(0.0153)	(0.0153)	(0.0211)
Chlamydia	0.000966	-0.000966	-0.000336	-0.00798	0.00288	0.000469	0.00150	-0.0103	-0.00287	-0.00544	-0.00424	-0.0167
	(0.00761)	(0.00763)	(0.00768)	(0.00946)	(0.00974)	(0.00973)	(0.00987)	(0.0130)	(0.0134)	(0.0135)	(0.0135)	(0.0168)
Consultation	-0.0973***	-0.101***	-0.0984***	-0.200***	-0.118***	-0.121***	-0.119***	-0.224***	-0.101***	-0.102***	-0.0990***	-0.199***
	(0.0286)	(0.0286)	(0.0280)	(0.0330)	(0.0321)	(0.0321)	(0.0318)	(0.0380)	(0.0339)	(0.0339)	(0.0336)	(0.0410)
					Panel B: Numbers of women using contraceptives							
Abortion	-0.000447	-0.000512	-0.000502	-0.000478	-0.000770**	-0.000894**	-0.000863**	-0.000857*	-0.000329	-0.000626	-0.000578	-0.000201
	(0.000305)	(0.000329)	(0.000338)	(0.000463)	(0.000366)	(0.000377)	(0.000369)	(0.000515)	(0.000490)	(0.000540)	(0.000539)	(0.000638)
Chlamydia	0.000630**	3.80e-06	5.56e-05	0.000528	0.000660	-3.84e-05	4.24e-05	0.00111*	0.000498	-0.000259	-0.000154	0.00104
	(0.000313)	(0.000493)	(0.000502)	(0.000407)	(0.000506)	(0.000780)	(0.000797)	(0.000675)	(0.000629)	(0.000932)	(0.000946)	(0.000759)
Consultation	-0.00146	-0.00271**	-0.00233**	-0.00424*	-0.00297**	-0.00388***	-0.00348**	-0.00556*	-0.00174	-0.00226	-0.00183	-0.00531**
	(0.00136)	(0.00121)	(0.00115)	(0.00244)	(0.00143)	(0.00149)	(0.00142)	(0.00285)	(0.00170)	(0.00168)	(0.00156)	(0.00269)
					Panel C: Log of numbers							
Abortion	-0.00204	-0.00250	-0.00247	-0.00311	-0.00319	-0.00399	-0.00387	-0.00547	-0.00170	-0.00357	-0.00336	-0.00273
	(0.00179)	(0.00192)	(0.00189)	(0.00254)	(0.00245)	(0.00262)	(0.00254)	(0.00340)	(0.00309)	(0.00338)	(0.00330)	(0.00439)
Chlamydia	0.00189	-0.00193	-0.00170	-0.000424	0.00116	-0.00326	-0.00291	0.000898	0.000327	-0.00448	-0.00397	-0.000169
	(0.00184)	(0.00204)	(0.00206)	(0.00275)	(0.00232)	(0.00264)	(0.00266)	(0.00394)	(0.00287)	(0.00344)	(0.00345)	(0.00465)
Consultation	-0.00807	-0.0160**	-0.0140*	-0.0339***	-0.0143*	-0.0203**	-0.0181**	-0.0390***	-0.00809	-0.0115	-0.00905	-0.0346***
	(0.00768)	(0.00758)	(0.00726)	(0.00833)	(0.00810)	(0.00838)	(0.00801)	(0.00991)	(0.00914)	(0.00972)	(0.00929)	(0.0105)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools	583	583	583	583	583	583	583	583	583	583	583	583
SCHOOL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
BACKGROUND	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
YEAR TREND	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES	NO
SCHOOL TREND	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression. The outcome is indicated in the rows. Each outcome should be read has having done or been treated for something before the end of the first (three first columns), second (three following columns) or third year of high-school (last three columns). We focus on the influence of non-Western women from the second and the third years of high school on non-Western immigrants from the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. In each of the three blocks of the table, the first column controls for school fixed effects. The second adds a linear trend to the first and the third adds individual controls to the second. The fourth column adds school specific linear trend to the third. Standard errors are clustered at the high school level.

Table 15: Influence of older native women on immigrant teenagers - Mechanisms

	Year 1				Year 2				Year 3			
Panel A: Fraction of usage												
Abortion	0.0131*	0.0136*	0.0129*	-0.000316	0.0137	0.0136	0.0128	-0.000414	0.0102	0.00663	0.00539	-0.00781
	(0.00725)	(0.00743)	(0.00735)	(0.00939)	(0.00959)	(0.00997)	(0.00973)	(0.0125)	(0.0111)	(0.0114)	(0.0111)	(0.0170)
Chlamydia	0.0135**	0.00661	0.00618	-0.00222	0.0163*	0.00921	0.00829	-0.00232	0.0177	0.00932	0.00739	-0.00618
	(0.00606)	(0.00594)	(0.00589)	(0.00940)	(0.00947)	(0.0100)	(0.00994)	(0.0124)	(0.0124)	(0.0125)	(0.0122)	(0.0168)
Consultation	0.0382	0.0249	0.0192	0.00688	0.0430	0.0362	0.0293	0.0185	0.0239	0.0215	0.0140	0.0158
	(0.0283)	(0.0301)	(0.0285)	(0.0347)	(0.0336)	(0.0351)	(0.0333)	(0.0439)	(0.0372)	(0.0386)	(0.0369)	(0.0502)
Panel B: Numbers of women using contraceptives												
Abortion	1.12e-05	1.37e-05	1.52e-05	5.72e-05	-7.89e-07	-4.72e-06	-1.05e-06	2.06e-05	-7.17e-06	-4.12e-05	-3.40e-05	2.49e-06
	(2.75e-05)	(3.13e-05)	(3.15e-05)	(6.13e-05)	(4.05e-05)	(4.54e-05)	(4.57e-05)	(7.77e-05)	(4.49e-05)	(5.09e-05)	(5.16e-05)	(9.33e-05)
Chlamydia	6.79e-05***	4.76e-06	1.05e-05	3.79e-05	8.97e-05***	2.47e-05	3.56e-05	0.000128	7.61e-05*	1.69e-06	2.00e-05	0.000112
	(2.38e-05)	(2.62e-05)	(2.71e-05)	(7.26e-05)	(3.18e-05)	(3.47e-05)	(3.58e-05)	(8.23e-05)	(4.31e-05)	(4.96e-05)	(5.02e-05)	(9.57e-05)
Consultation	3.02e-05	-9.08e-05	-3.91e-05	-0.000228	1.82e-05	-4.33e-05	1.88e-05	-0.000179	6.15e-05	4.50e-05	0.000115	-6.86e-05
	(0.000123)	(0.000129)	(0.000127)	(0.000187)	(0.000136)	(0.000149)	(0.000145)	(0.000212)	(0.000134)	(0.000151)	(0.000149)	(0.000221)
Panel C: Log of numbers												
Abortion	0.000124	0.000121	0.000129	-6.58e-05	-0.000245	-0.000415	-0.000347	-0.000583	-0.000322	-0.00130	-0.00112	-0.000972
	(0.000909)	(0.000996)	(0.000988)	(0.00132)	(0.00112)	(0.00123)	(0.00122)	(0.00166)	(0.00137)	(0.00154)	(0.00152)	(0.00204)
Chlamydia	0.00278***	0.00111	0.00126	0.00124	0.00315***	0.00132	0.00156	0.00261	0.00306**	0.00107	0.00144	0.00263
	(0.000736)	(0.000820)	(0.000833)	(0.00148)	(0.00111)	(0.00124)	(0.00125)	(0.00176)	(0.00145)	(0.00163)	(0.00163)	(0.00219)
Consultation	-0.000225	-0.00399	-0.00279	-0.00918**	-0.00297	-0.00569	-0.00427	-0.00835	-0.000646	-0.00188	-0.000225	-0.00279
	(0.00346)	(0.00355)	(0.00344)	(0.00420)	(0.00412)	(0.00446)	(0.00430)	(0.00557)	(0.00457)	(0.00501)	(0.00483)	(0.00602)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools	583	583	583	583	583	583	583	583	583	583	583	583
SCHOOL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
BACKGROUND	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
YEAR TREND	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES	NO
SCHOOL TREND	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression. The outcome is indicated in the rows. Each outcome should be read has having done or been treated for something before the end of the first (three first columns), second (three following columns) or third year of high-school (last three columns). We focus on the influence of native women from the second and the third years of high school on non-Western immigrants from the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. In each of the three blocks of the table, the first column controls for school fixed effects. The second adds a linear trend to the first and the third adds individual controls to the second. The fourth column adds school specific linear trend to the third. Standard errors are clustered at the high school level.

Table 16: Influence of older native women on native teenagers - Mechanisms

	Year 1				Year 2				Year 3			
Panel A: Fraction of usage												
Abortion	0.00274	0.000312	0.000377	-0.00175	0.00260	-0.00142	-0.00133	-0.00555	0.00142	-0.00398	-0.00386	-0.00848**
	(0.00228)	(0.00242)	(0.00239)	(0.00299)	(0.00280)	(0.00293)	(0.00290)	(0.00368)	(0.00341)	(0.00351)	(0.00349)	(0.00419)
Chlamydia	0.0294***	0.00171	0.00185	0.00403	0.0408***	0.00377	0.00408	-0.00294	0.0520***	0.00909	0.00945	0.00125
	(0.00526)	(0.00435)	(0.00432)	(0.00525)	(0.00723)	(0.00607)	(0.00601)	(0.00727)	(0.00872)	(0.00731)	(0.00723)	(0.00852)
Consultation	0.0758***	-0.00366	-0.00268	-0.0413***	0.0461***	-0.0142	-0.0134	-0.0502***	0.0299***	-0.00937	-0.00873	-0.0329***
	(0.0130)	(0.0110)	(0.0109)	(0.0142)	(0.0120)	(0.0109)	(0.0109)	(0.0135)	(0.0105)	(0.0101)	(0.0101)	(0.0125)
Panel B: Numbers of women using contraceptives												
Abortion	1.57e-05**	1.76e-06	-1.59e-08	-2.33e-05	1.67e-05**	-5.89e-06	-8.94e-06	-2.96e-05*	1.97e-05*	-9.56e-06	-1.42e-05	-2.71e-05
	(6.96e-06)	(7.82e-06)	(7.79e-06)	(1.65e-05)	(8.28e-06)	(9.25e-06)	(9.26e-06)	(1.66e-05)	(1.09e-05)	(1.24e-05)	(1.24e-05)	(2.37e-05)
Chlamydia	0.000109***	-4.81e-05***	-5.15e-05***	-4.48e-05*	0.000130***	-8.16e-05***	-8.66e-05***	3.96e-05	0.000136***	-0.000113***	-0.000118***	0.000107**
	(1.88e-05)	(1.86e-05)	(1.86e-05)	(2.57e-05)	(2.71e-05)	(2.80e-05)	(2.81e-05)	(4.42e-05)	(2.87e-05)	(2.97e-05)	(2.99e-05)	(5.00e-05)
Consultation	0.000260***	-0.000193***	-0.000207***	-0.000365***	0.000175***	-0.000166***	-0.000177***	-0.000264***	0.000141***	-7.97e-05**	-8.71e-05**	-0.000123**
	(3.60e-05)	(4.00e-05)	(3.91e-05)	(7.68e-05)	(3.45e-05)	(3.98e-05)	(3.94e-05)	(6.23e-05)	(3.27e-05)	(3.68e-05)	(3.65e-05)	(5.59e-05)
Panel C: Log of numbers												
Abortion	0.000475**	5.74e-05	-1.08e-05	-0.000185	0.000391	-0.000294	-0.000413	-0.000451	0.000370	-0.000528	-0.000711*	-0.000657
	(0.000239)	(0.000261)	(0.000259)	(0.000379)	(0.000282)	(0.000307)	(0.000305)	(0.000432)	(0.000349)	(0.000384)	(0.000384)	(0.000517)
Chlamydia	0.00366***	-0.000862	-0.000991*	0.000492	0.00401***	-0.00212***	-0.00231***	0.00130	0.00525***	-0.00172**	-0.00194**	0.00476***
	(0.000519)	(0.000537)	(0.000536)	(0.000631)	(0.000716)	(0.000740)	(0.000738)	(0.000895)	(0.000802)	(0.000831)	(0.000828)	(0.00108)
Consultation	0.00914***	-0.00375***	-0.00419***	-0.00337**	0.00613***	-0.00356***	-0.00387***	-0.00257*	0.00486***	-0.00145	-0.00167	-0.000122
	(0.00117)	(0.00124)	(0.00121)	(0.00170)	(0.00108)	(0.00114)	(0.00113)	(0.00156)	(0.00101)	(0.00105)	(0.00105)	(0.00145)
Nb of observations	201,264	201,264	201,264	201,264	201,264	201,264	201,264	201,264	201,264	201,264	201,264	201,264
Nb of schools	647	647	647	647	647	647	647	647	647	647	647	647
SCHOOL FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
BACKGROUND	NO	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
YEAR TREND	NO	YES	YES	NO	NO	YES	YES	NO	NO	YES	YES	NO
SCHOOL TREND	NO	NO	NO	YES	NO	NO	NO	YES	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression. The outcome is indicated in the rows. Each outcome should be read has having done or been treated for something before the end of the first (three first columns), second (three following columns) or third year of high-school (last three columns). We focus on the influence of native women from the second and the third years of high school on native teenagers from the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. In each of the three blocks of the table, the first column controls for school fixed effects. The second adds a linear trend to the first and the third adds individual controls to the second. The fourth column adds school specific linear trend to the third. Standard errors are clustered at the high school level.

Table 17: Influence of immigrant women (cohort) on immigrant teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	-0.00127	-0.00209	-0.0112**	-0.0112**	-0.0106**	-0.0156**
	(0.00478)	(0.00472)	(0.00499)	(0.00497)	(0.00501)	(0.00618)
Year 2	0.112***	0.100***	-0.107**	-0.109**	-0.0986**	-0.216***
	(0.0383)	(0.0353)	(0.0425)	(0.0426)	(0.0401)	(0.0468)
Year 3	0.124***	0.112***	-0.0831*	-0.0849*	-0.0744*	-0.179***
	(0.0404)	(0.0374)	(0.0439)	(0.0441)	(0.0417)	(0.0477)
			Panel B: Numbers of women using contraceptives			
Year 1	-0.00617**	-0.00448*	-0.0150***	-0.0160***	-0.0148***	-0.0315***
	(0.00310)	(0.00228)	(0.00409)	(0.00409)	(0.00404)	(0.00647)
Year 2	-0.00854**	-0.00574**	-0.0118**	-0.0127**	-0.0111**	-0.0241***
	(0.00372)	(0.00287)	(0.00506)	(0.00510)	(0.00498)	(0.00694)
Year 3	-0.00996**	-0.00620*	-0.0114**	-0.0123**	-0.0106**	-0.0237***
	(0.00503)	(0.00369)	(0.00533)	(0.00542)	(0.00530)	(0.00716)
			Panel C: Log of numbers			
Year 1	-0.0224**	-0.0166**	-0.0370***	-0.0396***	-0.0361***	-0.0700***
	(0.00917)	(0.00737)	(0.0133)	(0.0130)	(0.0128)	(0.0185)
Year 2	-0.0299***	-0.0209**	-0.0302*	-0.0328**	-0.0283*	-0.0558***
	(0.0116)	(0.00974)	(0.0166)	(0.0165)	(0.0161)	(0.0211)
Year 3	-0.0329**	-0.0210*	-0.0280*	-0.0305*	-0.0257	-0.0560***
	(0.0138)	(0.0110)	(0.0168)	(0.0168)	(0.0164)	(0.0207)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools			583	583	583	583
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of non-Western immigrants in the first year of high school on other non-Western immigrants in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives before the start of high school, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor. The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

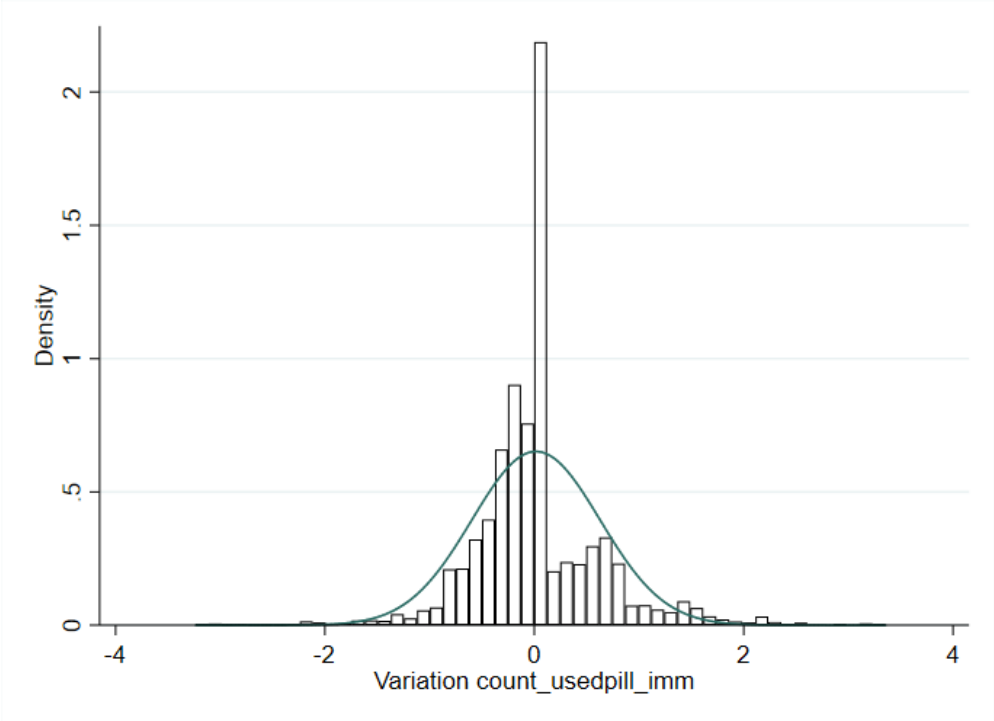
Table 18: Influence of native women (cohort) on immigrant teenagers - Contraceptives usage

			Panel A: Fraction of usage			
Year 1	0.184***	0.186***	0.0421	0.0207	0.0121	-0.0223
	(0.0305)	(0.0301)	(0.0436)	(0.0474)	(0.0463)	(0.0534)
Year 2	0.224***	0.226***	0.0750	0.0587	0.0463	0.0264
	(0.0366)	(0.0352)	(0.0503)	(0.0540)	(0.0523)	(0.0631)
Year 3	0.214***	0.218***	0.0506	0.0308	0.0147	-0.0168
	(0.0410)	(0.0384)	(0.0605)	(0.0653)	(0.0630)	(0.0724)
			Panel B: Numbers of women using contraceptives			
Year 1	0.00132***	0.00121***	0.000648	0.000426	0.000474	0.000266
	(0.000238)	(0.000228)	(0.000446)	(0.000465)	(0.000450)	(0.000569)
Year 2	0.00174***	0.00157***	0.00105*	0.000893	0.000923	0.000912
	(0.000308)	(0.000291)	(0.000563)	(0.000597)	(0.000573)	(0.000740)
Year 3	0.00175***	0.00154***	0.000880	0.000687	0.000698	0.000359
	(0.000310)	(0.000273)	(0.000541)	(0.000597)	(0.000566)	(0.000682)
			Panel C: Log of numbers			
Year 1	0.0393***	0.0352***	0.0183**	0.0123	0.00985	0.00828
	(0.00490)	(0.00466)	(0.00913)	(0.0101)	(0.00961)	(0.0118)
Year 2	0.0523***	0.0460***	0.0219*	0.0164	0.0127	0.0176
	(0.00597)	(0.00558)	(0.0112)	(0.0122)	(0.0116)	(0.0147)
Year 3	0.0540***	0.0461***	0.0135	0.00627	0.00152	0.00860
	(0.00738)	(0.00661)	(0.0127)	(0.0141)	(0.0137)	(0.0169)
Nb of observations	12,082	12,082	12,082	12,082	12,082	12,082
Nb of schools			583	583	583	583
SCHOOL FE	NO	NO	YES	YES	YES	YES
BACKGROUND	NO	YES	NO	NO	YES	YES
GRADE CHAR	NO	NO	NO	NO	NO	NO
YEAR TREND	YES	YES	NO	YES	YES	NO
SCHOOL FE*YEAR TREND	NO	NO	NO	NO	NO	YES

Note: Each cell of this table reports the coefficient of a separate regression where the outcome is having used contraceptives at least once by the end of the first, second and third years of high school (respectively rows labelled 'Year 1', 'Year 2' and 'Year 3'). We focus on the influence of native teenagers in the first year of high school on non-Western immigrants in the first year. In the first three rows, peers behaviour is measured as the fraction (of the influential population) of girls who already took contraceptives before the start of high school, in the following three rows the number of these girls and in the last three rows the log of the number. In the first three rows, the size of the group is added as a regressor.The first column controls for a year trend, while the second controls for a year trend and individual controls. The third column controls for school fixed effects. The fourth adds a year trend to the third and the fifth adds individual controls to the fourth. The sixth column adds school specific linear trend to the fifth. Standard errors are clustered at the high school level.

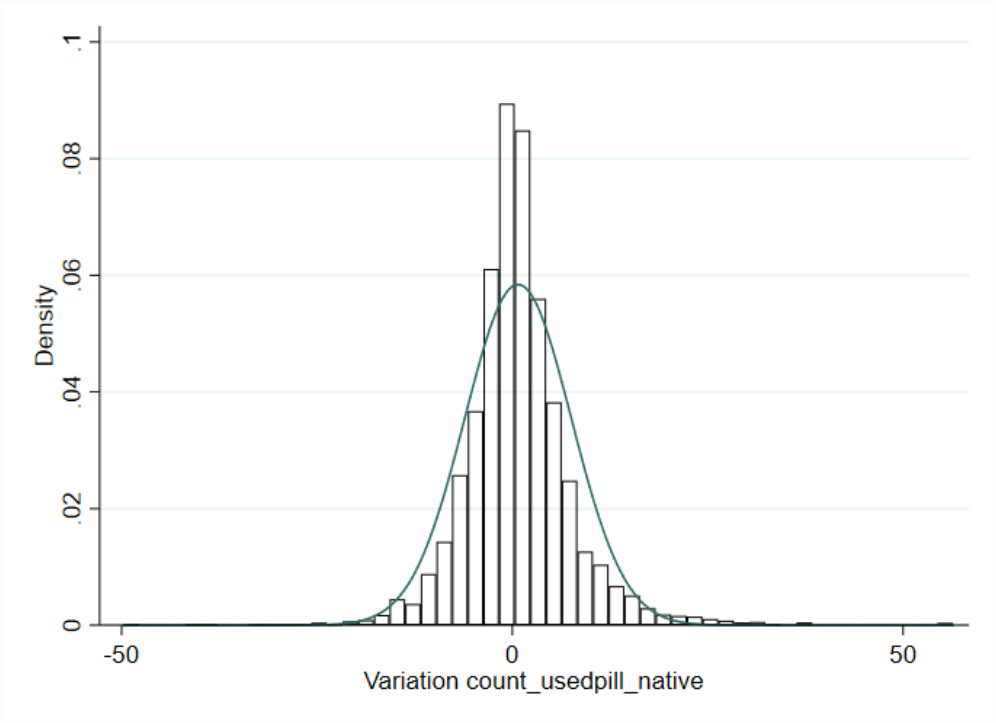
10 Appendix

Figure A1: Residual variation for identification (immigrant women from the same cohort)



Note : This graph plots the residuals of a regression of the number of immigrant women from the first year of high school (by school/cohort) who have taken contraceptives before the start of high-school on school fixed effects and a school specific time trend. The residual is the difference between the predicted and actual numbers (at the school/cohort level). The curve is a fitted normal distribution.

Figure A2: Residual variation for identification (native women from the same cohort)



Note : This graph plots the residuals of a regression of the number of native women from the first year of high school (by school/cohort) who have taken contraceptives before the start of high-school on school fixed effects and a school specific time trend. The residual is the difference between the predicted and actual numbers (at the school/cohort level). The curve is a fitted normal distribution.

Table A1: Members of the Organisation of Islamic Cooperation

Republic of AZERBAIJAN, Hashemite Kingdom of JORDAN, Islamic Republic of AFGHANISTAN
Republic of ALBANIA, State of The UNITED ARAB EMIRATES
Republic of INDONESIA, Republic of UZBEKISTAN, Republic of UGANDA
Islamic Republic of IRAN, Islamic Republic of PAKISTAN, Kingdom of BAHRAIN
BRUNEI-DARUSSALAM, People's Republic of BANGLADESH, Republic of BENIN
BURKINA-FASO, Republic of TAJIKISTAN, Republic of TURKEY
Turkmenistan, Republic of CHAD, Republic of TOGO, Republic of TUNISIA
People's Democratic Republic of ALGERIA, Republic of DJIBOUTI
Kingdom of SAUDI ARABIA, Republic of SENEGAL, Republic of The SUDAN
SYRIAN Arab Republic, Republic of SURINAME, Republic of SIERRA LEONE
Republic of SOMALIA, Republic of IRAQ, Sultanate of OMAN, Republic of GABON
Republic of The Gambia, Republic of GUYANA, Republic of GUINEA
Republic of GUINEA-BISSAU, State of PALESTINE, Union of The COMOROS
KYRGYZ Republic, State of QATAR, Republic of KAZAKHSTAN
Republic of CAMEROON, Republic of COTE D'IVOIRE, State of KUWAIT
Republic of LEBANON, Libya, Republic of MALDIVES, Republic of MALI, MALAYSIA
Arab Republic of EGYPT, Kingdom of MOROCCO, Islamic Republic of MAURITANIA
Republic of MOZAMBIQUE, Republic of NIGER, Federal Republic of NIGERIA
Republic of YEMEN

Table A2: Sample criteria - step-by-step selection

Women born 1981 to 1993 and in the country at age 13	385691
Entering high school in the first available grade - wide definition	270355
Entering high school in the first available grade - narrow definition	268258
Starting high school between the year they turn age 14 and 19	260243
Starting in July through September	258460
More than 10 people on the cohort	258248
More than 95% start in between the year they turn 14 to 19	213346

Note: This table reports the number of observations which satisfy the different criteria used to define the sample. These criteria are entered successively, each row being incremented after the one above itself.

Table A3: Balancing tests - Immigrants women

	Fraction	Log	Number	Fraction	Log	Number
	Immigrant women in older grades			Native women in older grades		
Months of education, mom	-3.86e-06	7.96e-05	0.000251	1.11e-05	0.000200*	0.00472
	(1.64e-05)	(5.80e-05)	(0.000187)	(1.24e-05)	(0.000108)	(0.00302)
Education above high school, mom	0.00121	0.00745	0.0544	0.000937	0.0397	0.833
	(0.00366)	(0.0115)	(0.0392)	(0.00315)	(0.0250)	(0.608)
Missing education, mom	9.79e-05	-0.0113	-0.0191	-0.00148	-0.0259	-0.451
	(0.00230)	(0.00815)	(0.0278)	(0.00183)	(0.0161)	(0.437)
Unemployed, mom	0.00293	-0.00145	0.0262	0.00195	0.0413*	1.085*
	(0.00412)	(0.0123)	(0.0433)	(0.00274)	(0.0212)	(0.569)
Out of the labour market, mom	0.00327	0.0117	0.0472	-0.00202	-0.0150	-0.433
	(0.00224)	(0.00732)	(0.0294)	(0.00181)	(0.0149)	(0.387)
Missing employment info, mom	-0.0107	-0.0721**	-0.366	0.0120*	-0.0167	-2.504
	(0.00738)	(0.0284)	(0.225)	(0.00629)	(0.0549)	(1.783)
Months of education, dad	5.59e-06	4.73e-05	0.000135	-9.25e-06	0.000113	0.00223
	(1.52e-05)	(5.30e-05)	(0.000171)	(1.07e-05)	(8.90e-05)	(0.00223)
Education above high school, dad	-0.00104	0.00341	0.0382	-0.00224	0.0193	0.381
	(0.00328)	(0.0100)	(0.0434)	(0.00241)	(0.0210)	(0.534)
Missing education, dad	-0.000841	-0.00608	-0.0145	0.000257	-0.0150	-0.168
	(0.00228)	(0.00816)	(0.0287)	(0.00178)	(0.0151)	(0.383)
Unemployed, dad	0.00392	0.0133	0.0914	0.00264	0.0585**	1.239
	(0.00412)	(0.0172)	(0.0918)	(0.00315)	(0.0254)	(0.802)
Out of the labour market, dad	0.00175	-0.000807	0.0174	-0.000235	-0.0215	-0.690*
	(0.00265)	(0.00750)	(0.0337)	(0.00200)	(0.0145)	(0.366)
Missing employment info, dad	-0.00716*	-0.0105	-0.0599	0.00306	-0.00493	-0.219
	(0.00390)	(0.0117)	(0.0512)	(0.00273)	(0.0234)	(0.609)
Age at birth, mom	5.13e-05	0.000558	0.00242	-0.000104	0.00126	0.0164
	(0.000188)	(0.000656)	(0.00271)	(0.000142)	(0.00112)	(0.0267)
Missing birth information, mom	-0.0164	-0.0788**	-0.409	0.00950	-0.0403	-2.634
	(0.0103)	(0.0354)	(0.316)	(0.00734)	(0.0727)	(2.165)
Age at birth, dad	8.10e-05	-0.000127	-0.00145	-5.66e-05	-6.42e-05	-0.00847
	(0.000111)	(0.000391)	(0.00165)	(8.76e-05)	(0.000638)	(0.0152)
Missing birth information, dad	-0.00594	-0.00577	-0.0211	0.00194	0.000408	-0.326
	(0.00453)	(0.0135)	(0.0498)	(0.00294)	(0.0254)	(0.692)
Parents married or cohabiting	-0.000122	0.00314	0.0108	-0.00437**	-0.0149	0.101
	(0.00254)	(0.00869)	(0.0433)	(0.00194)	(0.0164)	(0.455)
Nb observations	12,082	12,082	12,082	12,082	12,082	12,082

Note : Each cell of this table reports the coefficient of a separate regression. The outcome variable is indicated in the rows. The regressosrs are school fixed effects and a school specific linear time trend and a measure of peer behaviour. In the first three columns, the peers are the non-Western immigrant teenagers from older cohorts, while in the last three columns there are the native young women from older cohorts. Their behaviour is measured as a share of the influential population which has taken contraceptives at least once (column1 and 4), the number of women (numerator of the share, in columns 2 and 5) and the log of this number in columns 3 and 6. Observations consist of first year high school students with a non-Western background. Standard errors are clustered at the school level.

Table A4: Balancing tests - Native women

	Fraction	Log	Number	Fraction	Log	Number
	Immigrant women in older grades			Native women in older grade		
Months of education, mom	2.69e-05**	2.93e-05	0.000121**	1.22e-06	7.55e-05*	0.00401***
	(1.05e-05)	(1.82e-05)	(5.67e-05)	(5.08e-06)	(4.28e-05)	(0.00108)
Education above high school, mom	0.000415	-0.000783	-0.00305	0.000200	0.00140	0.148*
	(0.000738)	(0.00143)	(0.00521)	(0.000333)	(0.00299)	(0.0836)
Missing education, mom	-0.00437	-0.00812*	-0.0264*	3.80e-05	0.00516	-0.171
	(0.00273)	(0.00470)	(0.0153)	(0.00124)	(0.00979)	(0.279)
Unemployed, mom	-0.000972	0.00880**	0.0211	0.00103	0.0488***	0.970***
	(0.00230)	(0.00409)	(0.0129)	(0.00113)	(0.0100)	(0.241)
Out of the labour market, mom	0.00114	0.00132	-0.000619	-3.99e-05	0.0204***	0.387**
	(0.00135)	(0.00251)	(0.0103)	(0.000722)	(0.00600)	(0.151)
Missing employment info, mom	-0.000247	0.00242	0.0256	0.00106	0.00837	-0.0532
	(0.00414)	(0.00690)	(0.0272)	(0.00209)	(0.0164)	(0.413)
Months of education, dad	9.62e-06	1.31e-05	0.000110	4.99e-06	3.94e-05	0.00196**
	(7.82e-06)	(1.39e-05)	(8.22e-05)	(3.87e-06)	(3.34e-05)	(0.000878)
Education above high school, dad	0.000432	0.000443	0.00522	0.000848**	0.00802**	0.206**
	(0.000812)	(0.00144)	(0.00488)	(0.000423)	(0.00362)	(0.0969)
Missing education, dad	-0.00116	-0.00239	-0.0142	-0.000506	0.00385	-0.0513
	(0.00159)	(0.00274)	(0.0111)	(0.000803)	(0.00712)	(0.181)
Unemployed, dad	0.000187	0.000498	0.00809	-0.00210	0.0586***	1.001***
	(0.00244)	(0.00515)	(0.0184)	(0.00138)	(0.0114)	(0.282)
Out of the labour market, dad	0.000771	0.00374	-0.0112	-0.000591	0.00302	-0.0906
	(0.00181)	(0.00333)	(0.0176)	(0.000944)	(0.00841)	(0.218)
Missing employment info, dad	-0.000467	-0.00280	-0.0155	3.22e-05	0.00387	-0.0136
	(0.00207)	(0.00332)	(0.0108)	(0.000980)	(0.00812)	(0.198)
Age at birth, mom	1.57e-06	-0.000102	-0.000108	3.87e-05	-0.000679*	-0.0155
	(8.28e-05)	(0.000121)	(0.000367)	(4.51e-05)	(0.000365)	(0.00970)
Missing birth information, mom	-0.00128	0.0250	0.194	-0.00823	0.0598	3.726
	(0.0179)	(0.0447)	(0.310)	(0.00850)	(0.0697)	(2.334)
Age at birth, dad	-2.20e-06	0.000166	0.000433	3.19e-05	-0.000103	0.00297
	(5.85e-05)	(0.000110)	(0.000336)	(3.14e-05)	(0.000255)	(0.00628)
Missing birth information, dad	-0.00237	-0.0101*	-0.0350**	-0.00213	-0.0111	-0.520
	(0.00318)	(0.00560)	(0.0174)	(0.00183)	(0.0149)	(0.331)
Parents married or cohabiting	-0.000364	-0.00224	-0.00153	0.000734*	-0.000747	0.00603
	(0.000854)	(0.00162)	(0.00511)	(0.000430)	(0.00354)	(0.0995)
Nb observations	201,264	201,264	201,264	201,264	201,264	201,264

Note : Each cell of this table reports the coefficient of a separate regression. The outcome variable is indicated in the rows. The regressosrs are school fixed effects and a school specific linear time trend and a measure of peer behaviour. In the first three columns, the peers are the non-Western immigrant teenagers from older cohorts, while in the last three columns there are the native young women from older cohorts. Their behaviour is measured as a share of the influential population which has taken contraceptives at least once (column1 and 4), the number of women (numerator of the share, in columns 2 and 5) and the log of this number in columns 3 and 6. Observations consist of native first year high school students. Standard errors are clustered at the school level.

Table A5: Balancing tests - Measures of concentration

	Fraction	Log	Number	Fraction	Log	Number
	Immigrant Women			Native Women		
Months of education, mom	-1.08e-05	-0.000104**	-0.000610	-3.64e-06**	-5.40e-05**	-0.000211
	(7.14e-06)	(5.17e-05)	(0.000610)	(1.61e-06)	(2.66e-05)	(0.000133)
Education above high school, mom	-0.000577	-0.0120	-0.0678	-0.000346***	-0.00440**	-0.0196**
	(0.00117)	(0.0110)	(0.102)	(0.000111)	(0.00178)	(0.00965)
Missing education, mom	0.00189*	0.0145*	0.127	0.000312	0.0110*	0.0368
	(0.00111)	(0.00753)	(0.0824)	(0.000420)	(0.00651)	(0.0308)
Unemployed, mom	0.000218	-0.000919	0.0792	0.000129	-0.000988	0.0133
	(0.00164)	(0.0115)	(0.105)	(0.000377)	(0.00555)	(0.0257)
Out of the labour market, mom	0.000908	0.00208	0.0190	0.000410*	0.00540	0.00658
	(0.000759)	(0.00754)	(0.0727)	(0.000227)	(0.00330)	(0.0177)
Missing employment info, mom	-0.00225	-0.00816	-0.331	0.000164	0.00782	0.0886
	(0.00303)	(0.0258)	(0.298)	(0.000778)	(0.0112)	(0.0546)
Months of education, dad	-2.18e-06	-8.54e-05*	-0.000480	-1.51e-06	-2.71e-05	-4.94e-08
	(5.58e-06)	(4.52e-05)	(0.000375)	(1.33e-06)	(2.03e-05)	(0.000112)
Education above high school, dad	-0.00115	-0.0177*	-0.175**	-0.000183	-0.00335	-0.0146
	(0.000898)	(0.0104)	(0.0752)	(0.000128)	(0.00205)	(0.0108)
Missing education, dad	-3.00e-05	0.0112	0.0357	0.000122	0.00239	0.00440
	(0.000954)	(0.00724)	(0.0635)	(0.000263)	(0.00401)	(0.0237)
Unemployed, dad	-0.000918	0.00971	0.127	0.000569	0.00271	0.0240
	(0.00322)	(0.0145)	(0.168)	(0.000423)	(0.00589)	(0.0341)
Out of the labour market, dad	-0.000150	-0.00574	-0.0337	8.56e-05	0.00682	0.0203
	(0.000887)	(0.00766)	(0.0724)	(0.000306)	(0.00446)	(0.0233)
Missing employment info, dad	-0.000763	0.00582	-0.167*	-0.000313	-0.00187	-0.0279
	(0.00123)	(0.0109)	(0.0961)	(0.000336)	(0.00486)	(0.0262)
Age at birth, mom	-2.74e-05	-0.000200	-0.00189	-1.39e-05	-0.000132	-0.000449
	(9.59e-05)	(0.000597)	(0.00653)	(1.13e-05)	(0.000178)	(0.000906)
Missing birth information, mom	-0.00148	0.00218	-0.455	0.00132	0.0260	0.291
	(0.00297)	(0.0278)	(0.362)	(0.00379)	(0.0445)	(0.217)
Age at birth, dad	-4.26e-05	-0.000430	-0.00114	-1.18e-05	-0.000131	-0.000911
	(4.57e-05)	(0.000328)	(0.00340)	(8.44e-06)	(0.000129)	(0.000659)
Missing birth information, dad	0.000519	0.0203	-0.0498	0.000165	0.00851	0.0171
	(0.00138)	(0.0133)	(0.111)	(0.000624)	(0.00784)	(0.0414)
Parents married or cohabiting	0.00161*	0.00528	0.182*	-4.87e-05	-0.000787	0.00106
	(0.000901)	(0.00802)	(0.0956)	(0.000118)	(0.00198)	(0.0106)
Nb observations	12,082	12,082	12,082	201,264	201,264	201,264

Note : Each cell of this table reports the coefficient of a separate regression. The outcome variable is indicated in the rows. The regressosrs are school fixed effects and a school specific linear time trend and a measure of peer composition. Composition is measured as the share of non-Western immigrants in the cohort (column1 and 4), the number of women with a non-Western background (numerator of the share, in columns 2 and 5) and the log of this number in columns 3 and 6. Observations consist of first year high school students (from the same cohort) with a non-Western background in the first three columns and native students in the last three columns. Standard errors are clustered at the school level.

Chapter 3

The Intergenerational (Im)mobility of Immigrants

1 Introduction

The study of the economic assimilation of immigrants is usually sliced into three non-overlapping periods. The first one is the transition from origin to destination country which analyzes how human capital can be transferred between labor markets (Friedberg (2000)). The second one focuses on the dynamics of earnings once in the immigration country and deals with how to accumulate back human capital (Eckstein and Weiss (2004), Cohen-Goldner and Eckstein (2008)). The third one is the intergenerational mobility of immigrants that tries to understand how human capital is transmitted across generations (Borjas (1993), Aydemir et al. (2009)). In this paper, I show that these different periods should be considered together. They are all part of the same story.

Consider the following hypothetical example: an intermediate level white-collar from a low-income country migrates to Europe. For many reasons (he does not write well in the language of the country, he cannot rely on his professional network...), he cannot find a similar job than the one he had before migration and ends up working as a construction worker. In the destination country, he is indistinguishable from another construction worker who was an unskilled agricultural worker before migration. Yet they are maybe very different and their children may have different perspectives.

Looking separately at the three periods can be misleading because migration is a downward shock for many immigrants. At the individual level, it brings a discontinuity in family social histories. At the aggregate level, it compresses pre-existing differences. The social statuses of immigrants are more homogeneous in the country of immigration than they were in the country of origin. Omitting to take into account pre-migration information can bias upwards the estimates of social mobility between the first and second generations. Omitting to consider the loss of human

capital (HC) at migration takes out of the picture a crucial step of the dynamics of economic assimilation. Omitting the situation of second generation immigrants biases downwards the estimates of assimilation as part of it occurs through them.

To cover these three periods, one needs to extend the usual time bounds of the analysis. To do so, I use unique French survey data to trace family histories over three generations partly in the sending country, partly in France. To identify the economic backgrounds at different points in time, I rely on information about occupation and education of first and second generation immigrants. For first generation (and their parents, the grandparents from the perspective of the second generation), this information is available for the period before they migrated and in the country of destination together with the educational achievements of their children (who grew up in the country of immigration).

The data allows to go beyond descriptive evidence and dig into mechanisms. It contains information on the transition between labor markets (first job in France, investment into learning French...) and parental investment in children (schooling decision, help with homework...). I use it to contrast the process by which parents transfer their human capital between countries from the one by which they transmit it across generations.

In the first part of the paper, I test whether the heterogeneity in economic conditions over time has a U-shape. This corresponds to a situation in which the heterogeneity that existed in the country of origin was compressed at the time of migration and reappeared after a few years partly with the first generation but most noticeably with the second generation.

There is clear evidence of such a pattern. Around two thirds of the immigrants who had a relatively high socio-economic status (SES) in the country of origin end up in low-skilled occupations when migrating. There is very little mobility after migration as almost 90% keep the same type of occupations. Pre-migration status is a key determinant of the success of second generation immigrants. In a regression explaining the probability of finishing high school, the coefficient associated with the social background in the country of origin is of similar magnitude as the one associated with father's occupation in the destination country. These results are very robust as is documented at length in the online appendix.

In the second part, I formalize the intuition into a simple model, built around the idea that immigrants cannot entirely transfer their human capital between countries but can transmit it to their children. In the hypothetical example, it would mean that the intermediate level white collar who became a construction worker still raises his children as a white collar would. In the model, parents decide how much to invest into accumulating back human capital and how much into helping their children. To explain the U-pattern, I focus on two alternative scenarios: is it that parents invest less in themselves than in their children, i.e. is it driven by quantity? Or is it that investments in children pay more, i.e. is it driven by productivity? Is it that the cabdriver does not try to learn the language of the destination country but helps his children with their homework? Or is it that he does both but gains little from learning the language when his children

benefit from his help?

I find empirical evidence supporting the second scenario. First generation immigrants have difficulties accumulating back human capital after the initial loss of migration. Their returns to learning French are low as are their returns to experience in France. However, they did not suffer from the same downward shock when it comes to transmitting HC to their children. While on the labor market part of their parents' HC was lost, at home the children can fully benefit from it. Returns to investment in children (helping with homework, elaborating a schooling strategy ...) are high making it easier for second generation immigrants to accumulate HC.

The rest of the paper is organised as follows: section 2 reviews the literature and section 3 presents the data. Section 4 shows the descriptive evidence while section 5 outlines a model that helps understanding them. Section 6 tests empirically additional predictions of the model. Section 7 concludes.

2 Literature review

The first strand of literature this paper contributes to is the one on the economic assimilation of immigrants. Most papers focus on one (maximum two) of the three periods that I am combining. Seminal papers by Borjas (1995, 2015) have focused on the convergence in earnings with natives and how it differs across cohorts of immigrants. I address another research question since I compare immigrants with similar characteristics in the destination country and different pre-migration background. Another series of papers have used pre-migration situation to study the integration of immigrants in Israel. For instance, Friedberg (2000) showed that HC acquired pre-migration is not fully transferable to the destination country's labor market. Eckstein and Weiss (2004) and Cohen-Goldner and Eckstein (2008) estimate structural models of first generation earnings that include pre-migration characteristics. Their conclusion is that immigrants with higher HC at the time of migration keep this advantage once in the destination country. An important contribution with respect to these papers (besides the inclusion of information on the second generation) is that I look at a much larger sample. They follow around 500 high skilled migrants from the Soviet Union for which they have pre-migration information on occupation. In comparison, I use a representative sample of immigrants to a major European country and observe outcomes many years after migration (when Cohen-Goldner and Eckstein (2008) followed immigrants for 5 years). Previous studies have used more extensively information on pre-migration background but mostly in order to explain the decision to migrate (Chiquiar and Hanson (2005), Kaestner and Malamud (2014)) that I take as given in this paper.

My perspective is complementary to that of papers that look at the economic assimilation and educational achievements of second generation immigrants (as for instance in Algan et al. (2010), Card (2005), Gang and Zimmermann (2000)). I do not compare immigrants with natives as I am focusing on the heterogeneity among second generation immigrants. My results are different to that

of Gang and Zimmermann (2000) who find no effect of parental education on the achievements of second generation immigrants. The authors say in conclusion of their paper that parental education is not necessarily a good proxy of immigrants HC. In this paper, I can measure it through different dimensions.

Closest in spirit is Caponi (2011) who develops a dynamic model where first generation are ready to suffer a loss for the benefit of the second generation. The author matches survey data from the US and Mexico to estimate a structural model of immigrants earnings. I add to this analysis the possibility for parents to invest in adapting their HC and developing that of their children. I also use data which directly follows the same families before and after migration. This allows to better apprehend the (dis)continuities brought by migration at the individual level.

The second strand of literature this paper contributes to is the one on the intergenerational mobility of immigrants. Key papers in this literature include Borjas (1993) and Aydemir et al. (2009). They relate outcomes of first and second generation immigrants in the destination country controlling for country of origin. These papers do not include family specific information on pre-migration status. Origin country dummies may not capture well individual situations if immigrants from the same country are heterogeneous. Compared to these papers, I indirectly test whether the relevant background for assessing mobility is the one in the country of destination or the origin country. If it is the former, using the latter corresponds to a measurement error which can lead to a downward bias.

This paper also contributes to the general discussion on long term social mobility (see for instance Chan and Bolivier (2013); Mare (2011); Braun and Stuhler (2018)). This literature typically finds less mobility across multiple generations than was first thought by Becker and Tomes (1986). It is an open question whether it is driven by the presence of older generations (typically grandparents) or by the transmission of latent unobservable characteristics (see for instance Clark (2014)). This paper exploits a setting in which presence of older generations does not play a role as I focus on families where grandparents did not migrate. This paper thus provides evidence supporting the latter explanation.

To a lesser extent, this paper contributes to the "epidemiological" approach literature (for instance Fogli and Fernandez (2006); Fernandez (2007); Blau and Kahn (2015)). These papers have shown how origin country characteristics explain key outcomes of second generation immigrants, such as fertility or female labour force participation. I add a within origin country dimension of heterogeneity to this analysis.

This paper builds on a stylized fact already studied in sociology by Ichou (2014) namely that pre-migration relative educational achievements of parents matter for children. It emphasizes the analysis on HC transmission. In comparison, I add a dimension to the analysis, the portability of parents' HC, and therefore a step in the story, the transition into the labor market for the first generation. I also provide a theoretical framework to understand the descriptive evidence together with empirical tests of potential mechanisms.

3 Description of the data

3.1 One dataset, two samples

The dataset used in this paper is "Trajectoires et Origines" (TeO), collected by the National Institute of Demographic Studies (INED) and the National Statistical Agency (INSEE) in 2008/2009. It is a cross sectional survey (18,864 persons interviewed in total) based on representative samples of immigrants and natives. TeO has several advantages over alternative data sources such as labor force surveys and censuses. First, TeO has been designed with the specific purpose of studying first and second generation immigrants; it asks questions targeted for this population. Second, TeO is the only dataset in France to have specifically sampled 2nd generation immigrants. People do not appear as immigrants in civil registries if they are born in France. Second generation immigrants had to be sampled from registries of first generation immigrants. Last but not least, TeO contains information on the socio-economic status in the country of origin.

Three different populations are surveyed in TeO: (i) first generation immigrants, (ii) second generation immigrants and (iii) natives. The questions in the survey refer to different moments in the lives of the people being interviewed: the personal and family history, the current situation and the situation of the children. Table 1 shows how different sections of the questionnaire are combined to generate two samples of second generation immigrants. By relying on two samples instead of one, I use all the information available in the survey. I can also check that my results are consistent across samples. I do not match the two surveys and always use them separately.

[Table 1 here]

The sample that surveys first generation, from now on referred to as *sample first*, contains mostly information on the situation prior to migration and on the parents situation in France but less on the situation of the children. The sample that surveys second generation, from now on referred to as *sample second*, contains mostly information on the situation of the children and on parental investments in their children but less on the parents situation both in France and prior to migration. When information in both samples do not overlap, I use the most relevant one. When it does, I use the two samples to perform "out of sample" checks within TeO. Figures 1 and 2 show show a digram (respectively for *sample first* and *sample second*) of who is being interviewed and how they relate to the population of interest. In everything that follows, second generation immigrants refer to people whose both parents are immigrants. I restrict the analysis to families where children grew up in France (were born there or arrived before age 10) and parents accumulated their HC in the country of origin (had some professional experience or completed their education before migrating).

[Figures 1 and 2 here]

3.2 Definition of Socio-Economic Statuses

The notation works as follows: S stands for status, subscripts P and C stand for parents and children. There are three periods, $t = 0$ refers to pre-migration, $t = 1$ refers to the time of arrival in France, $t = 2$ refers to the situation at the time of the survey. Parents status is observed three times $S_{P,0}, S_{P,1}, S_{P,2}$ and children status only once, at $t = 2$, $S_{C,2}$. Status can take two values H or L, for high and low. There are therefore three categories of status that must be defined: the achievements of the second generation ($S_{C,2}$), the socio-economic situation (SES) of the parents in France ($S_{P,1}, S_{P,2}$) and in the country of origin ($S_{P,0}$).

It is relatively straightforward to define $S_{C,2}, S_{P,1}$ and $S_{P,2}$ since they all refer to the French environment. $S_{P,1}$ and $S_{P,2}$ are based on father's occupation[1]: they are considered "high" if he is a white collar, more specifically if he has a "high managerial" or "supervisory occupations"[2]. $S_{C,2}$ is the outcome variable (different for each sample) that captures the educational and labor market achievements of the children. For *sample first* it is defined as "high" for having obtained the baccalauréat (final exam in high school and most relevant outcome in this sample). For *sample second* it is separately defined as "high" for not having dropped out[3], having obtained the baccalauréat, having a higher education degree and being a white collar (same categories as for parents).

It is more difficult to establish a criterium for having a "high" status in the country of origin ($S_{P,0}$). Ideally, this definition would be country and time specific. Having a high school diploma means something different in a developed and in a developing country. It means something else for someone who grew up in the 70s than for someone who grew up in the 40s[4]. However, in this paper I use a binary variable based on definitions that are common to all countries and time periods. Since socio-economic status has multiple components, I include, when possible, two of its dimensions in the definition: occupation and education. Having information on two generations in *sample first*, parents and grandparents, I use two definitions, the *parents* and the *grandparents* definition.

According to the former, the family of a second generation immigrant is considered "high" in the country of origin if the parent had a "high" occupation before migrating or finished secondary education. According to the latter, a "high" household has one grandfather who had a "high" occupation or one grandparent who had above primary school level education. In *sample second*, the family of a second generation immigrant is considered "high" in the country of origin if the

[1] I focus on fathers' and not mothers' occupation because of potential biases due to female labor force participation.

[2] If the mother is being interviewed, I use information on the occupation of the partner provided they report the partnership having begun before the birth of the child whose observations are kept in the sample.

[3] Defined as having successfully completed junior high school, passing the exam known as the "Brevêt des Collèges".

[4] For instance, Ichou (2014) constructs a relative measure of parental educational achievements in the country of origins using TeO together with the data collected by Barro and Lee (2013).

father finished secondary education (to keep the definition in line with that of *sample first*).

Criteria developed for *sample first* have two advantages: they include several dimensions of the pre-migration background and use information far back in time from the perspective of second generation immigrants, capturing well long term persistence. For these reasons, it is my preferred sample to show the resurgence of family socio-economic status. Because the paper is focused on the transfer and the transmission of human capital, I use in my main specifications, the *parents* definitions as grandparents did not migrate.

By applying the same binary criteria to populations from different countries and time periods, I will compare different parts of the distribution for different subsamples. My objective is not to create a relative measure by countries but to answer the question: do those that were on top remain above the others, even after having fallen off the ladder? Since I use binary definitions, I compare different groups. What ultimately matters is that, in each country and each period, an important share of the sample belongs to each group.

In what follows the population of immigrants is divided in four groups: families with (i) $S_{P,0} = H, S_{P,2} = H$, the *always high*, (ii) $S_{P,0} = L, S_{P,2} = L$, the *always low*, (iii) $S_{P,0} = L, S_{P,2} = H$, the *upgraded*, (iv) $S_{P,0} = H, S_{P,2} = L$, the *downgraded*. The online appendix includes a more detailed presentation of the data, definition of the samples and of the statuses.

3.3 Descriptive Statistics

Table 2 provides basic descriptive statistics on the first generation immigrants who have adult children in *sample first*. I report the gender of the person being interviewed together with the distribution of their educational level and that of their fathers (the grandfathers from the perspective of second generation immigrants), the year in which they arrived and their occupation before migrating. Immigrants in the sample arrived at relatively similar periods since the 25^{th} and 75^{th} percentiles of the distribution of years of arrival are only nine years apart. There is however variation in their educational achievements, roughly three equal groups: below primary school, primary school and secondary school and above[5]. Among immigrants reporting pre-migration professional experience, more than 25% had a "high" occupation; showing this population had heterogenous backgrounds.

[Table 2 here]

Table 3 describes the percentages of immigrants coming from a “high” or “low” SES by sending country (or region of the World)[6] in *sample first* according to the *parents definition*. There is

[5]In the table A7 in the online appendix, a similar table is produced for the parents of immigrants followed in *sample second* (with the information available for that sample). The two samples are very different and correspond to different waves of migration (as can be seen from the distributions of years of arrival).

[6]The version of TeO that I use does not contain information on countries of origin for certain countries but rather group them by region.

always a substantial fraction of immigrants coming from a different SES than the one commonly associated with their country of origin. Not all immigrants from developed countries are labelled "high" and not all immigrants from developing countries are labelled "low". For example, 29% of the immigrants from Maghreb are from a “high” SES while Spain and Italy provide a large share (more than 70%) of “low” status immigrants. The situation for African countries is very heterogeneous, with countries such as Mali and Senegal providing mostly “low” SES immigrants and countries such as Ivory Cost or Congo Brazzaville providing more or less equal shares. What is important for this paper is that there are substantial pre-migration differences within countries.

[Table 3 here]

One of the contribution of this paper is to break down the immigrant population according to criteria at the junction of their status in the country of origin and in the country of destination. Table 4 reports descriptive statistics separately for immigrants from Europe, Asia, Maghreb and Africa and for (1) *always high*, (2) *always low* and (3) *downgraded*. I use data from *sample first* as it contains more information on pre-migration characteristics and report the mean year of arrival, the proportion of male, the proportion of people who reported good or very good level of (spoken) French when they arrived, the proportion of immigrants who completed secondary and higher education. I also report the mode(s) of occupation of the grandfather, the mode(s) of occupations of fathers and mothers in France.

[Table 4 here]

Table 4 provides an opportunity to check that combining simple definition of statuses can identify a group of *donwgraded* immigrants. When focusing on characteristics not used in the definitions of $S_{P,0}$ and $S_{P,2}$ *always low* and *downgraded* have a similar situation in France but a different one in the country of origin. They look the same after migration, they were different before. The mode of occupations in France[7] is similar for fathers and mothers when broken down by regions of origin (for men in all cases and for women from Europe and Africa). However the pre-migration characteristics of these two groups are different. For instance, *downgraded* immigrants from Maghreb and Sub-Saharan Africa are 78% and 85% to report being fluent in (spoken) French at the time of arrival when only 42% and 62% of the *always low* do. The grandfathers' occupation also differ between the groups (for Maghreb and Europe). On the other hand, the *downgraded* and the *always high* are different in most dimensions in France and in the country of origin.

[7] In table 4, I use more detailed information on occupation, namely a two digit classification when I only rely on the first digit in the definitions of $S_{P,0}$, $S_{P,1}$ and $S_{P,2}$

4 Resurgence of pre-migration characteristics

4.1 Compression of socio-economic heterogeneity following migration

Migration is a permanent negative shock for a large share of immigrants. To show that the shock is negative, I document how the heterogeneity in SES that existed in the country of origin was reduced at the time of migration. To show that the shock is permanent, I document the mobility in SES between the year of arrival in France and the time of the interview. I use observations from *sample first* since they are the only ones to contain information on SES at the three relevant periods, before migration, at arrival and now.

[Table 5 here]

Table 5 shows the transition matrix between the pre migration status and the status upon arrival for men[8]. The rows refer to the status in the country of origin, the columns to the status in France at time of arrival. Not surprisingly, almost all (97.14%) of those who had a low status prior to migration remain in that category once in France. What is more interesting is that a large share (63%), almost two thirds, who came from a "high" status before migration loses it.

This result is very robust. In the online appendix, I check whether it holds for different subsamples. In particular, I test the other definitions of "high" and "low" used in this paper (table A.8 to A.10). I also look at various regions of origin, namely Europe, Asia, Africa and Maghreb, relying on the *parents definition* in tables A.12 to A.15 and on the *grandparents definition* in tables A.16 to A.19. Although the probability of being downgraded is lower for Europeans, it is still relatively high (roughly 50%). I also look at various waves of migration by splitting the sample according to year of arrival in France (between 1950 and 1970, 1970 and 1990 and 1990 and 2010). Tables A.20 to A.22 show results for the *parents definition*, while tables A.23 to A.25 show results for the *grandparents definition*. The results are not driven by a specific period. The results hold in the two samples. Table A.11 reports the transition between $S_{P,0}$ and $S_{P,2}$ in *sample second*[9].

[Table 6 here]

Table 6 is a transition matrix where rows refer to status at the time of arrival and columns to status at the time of the interview. It shows the low level of mobility among first generation immigrants: the diagonal elements are above 85%. Mobility is not particularly higher when one

[8] I cannot construct a good measure of $S_{P,1}$ when the person being interviewed is a woman. To construct $S_{P,2}$ when the person being interviewed is a woman, I use information on the occupation of the partner provided they report the partnership having begun before the birth of the child whose observations are kept in the sample. TeO contains current occupation (of the person being interviewed and of the partner), however it does not have information on first occupation in France of the partner. For that reason, I limit the sample to children whose father is being interviewed in tables 5, 6 and 7.

[9] The proportion of *downgraded* is 51.64%, very similar to the same number for *sample first*, 52.82% from table 7.

focuses on immigrants with a previously "high" status. This can be seen in table 7 which produces the same transition matrix but only for those with $S_{P,0}$=H. The left diagonal element is lower in magnitude than in table 6, 75.8% instead of 89.74%, but still very large: three out of four *downgraded* immigrants remain in a low status over the years. These results are robust to similar sample restrictions than those mentioned above (see online appendix, tables A.26 to A.45).

[Table 7 here]

The population of first generation *downgraded* has difficulties being distinguished in the destination country from the population of *always low*. What made these populations different in the country of origin (level of education, abilities, social and professional networks...) fails to have been identified or to have operated once in the destination country. For comparison with natives, the last row of table 6 reports marginal distributions for non immigrants.

4.2 Reappearance of socio-economic heterogeneity with the second generation

To observe the importance of pre-migration characteristics, I estimate the following regression in *sample first*:

$$y_{i,f} = \alpha + \beta_1 1\left\{(S_{P,0})_{i,f} = H\right\} + \beta_2 1\left\{(S_{P,2})_{i,f} = H\right\} + \gamma X_i + \theta Z_f + \varepsilon_{i,f} \tag{3.1}$$

Since I observe potentially several children per parent, subscript i refers to children and f to families. $1\left\{(S_{P,0})_{i,f} = H\right\}$ is an indicator function taking value 1 when pre-migration status is "high" and 0 otherwise. X is a vector of controls at the individual level, Z at the family level. In the baseline specification, the controls at the individual level are age and gender and at the family level country of origin of the parent (to focus on the variation in status within countries). Standard errors are clustered at the family level.

The estimates for *sample first* are reported in the first three columns of table 8. In this sample, y is a dummy variable taking value one for having obtained the baccalaureat. The first column uses $S_{P,2}$ to capture family socio-economic situation in the destination country while the second column uses $S_{P,1}$ and the third uses the amount of monetary resources available to the household. The last measure accounts for measurement error that could arise from only focusing on fathers and from misclassification of occupations. For instance, many immigrants are self-employed; they are classified as "low" in my categorization (to keep it in line with the classification for natives), while potentially having high earnings.

[Table 8 here]

The additional probability of obtaining the baccalaureat associated with a "high" pre-migration background ranges from 12.9 and 15.7 percentage points. This is very large when one compares this

coefficient with the unconditional mean, it corresponds to a 20% to 25% increase in educational achievements. It is also large when compared to the coefficient associated with current family status, either $S_{P,1}$ or $S_{P,2}$. The advantage associated with coming from a "high" status pre-migration is within the same order of magnitude to that associated with having a white collar father in France.

It is interesting to note that the first job after migration is only marginally significant in explaining children achievements and that its associated coefficient is lower than the one of current occupation (8.9 vs 13.6 percentage points). It shows that the situation in the years that follow migration may not capture well the HC level of immigrants.

The results are confirmed in *sample second* for several additional outcome variables, respectively not dropping out of school, obtaining the baccalaureat, obtaining a higher education degree, log wage and having a 'high' occupation (according to the definition of $S_{P,1}$ and $S_{P,2}$). I report the results of estimating equation 3.1 in the upper part of table 9. The controls are the same than for *sample first* but regressions on labor market outcomes include age squared as a control and are limited to male children (due to selection into employment for female immigrants).

For all outcomes, I find a large effect of pre-migration characteristics, although not precisely estimated for log of wages. The advantages of having parents from a higher social background in the country of origin is not limited to finishing high school. Results appear particularly large for outcomes where the unconditional probability is low for immigrants, namely obtaining a higher education degree and having a 'high' occupation. Similar to the results in table 8, the coefficient associated with $S_{P,0} = H$ is within the same order of magnitude than the one associated with $S_{P,2} = H$.

For each regression, I report the number of observations whose parents are classified as "high" and "low" in the country of origin. This gives an idea of how the population of "low" background immigrants in France split between pre-migration status. In *sample first*, between 36 and 38% of the second generation immigrants come from a family with a "high" status (according to the *parents definition*) before migration[10]. The proportion is much lower in *sample second*, between 7 and 10% depending on the outcome used. First generation immigrants in this sample are older and were raised at times where average educational achievements were lower. The threshold of finishing secondary education is high for this population.

To see how the effect decomposes between *downgraded* and *always high*, I then estimate the following equation:

$$y_{i,f} = \alpha + \sum_{j=L,H} \sum_{l=L,H} \beta_{j,l} 1\left\{(S_{P,0})_{i,f} = j, (S_{P,2})_{i,f} = l\right\} + \gamma X_i + \theta Z_f + \varepsilon_{i,f} \tag{3.2}$$

where $1\left\{(S_{P,0})_{i,f} = H, (S_{P,2})_{i,f} = L\right\}$ is for instance an indicator function taking value one if

[10]The proportion is similar when using *grandparents definition*, between 30 and 35% as can be seen from table A.52 from the online appendix

an individual belongs to the group of *downgraded*. The results for *sample first* are reported in the last two columns of table 8. The difference in probability between *downgraded* and *always low*, 11.1 and 9.9 percentage points when measured with $S_{P,1}$ and $S_{P,2}$, is indicative that the U-shape representation is accurate. Indeed, those who were "high" are much more likely to get back to a higher status.

[Table 9 here]

Results for *sample second* are reported in the lower part of table 9. The same pattern emerges than for *sample first*. There is a 29 percentage points difference between *downgraded* and *always low* in the probability of having a higher education degree and a "high" occupation. Children of immigrants from a "low" SES in France who have a higher education degree or are white collars disproportionately come from a "high" strata in the country of origin. They retrieve the fallen position of their parents.

These results are very robust to varying samples and including powerful controls. First of all, results are similar when immigrants are classified according to the *Granparents* definition (see table A.52 in the online appendix). Also results do not change when one includes (for both samples) the following controls: number of siblings (for resources per capita) in tables A.46, A.53 and A.59, dummies for parents' year of birth in tables A.47, A.54 and A.60, dummies for the year of arrival in France (for specific waves of migration) in tables A.48, A.55 and A.61, dummies for living in an ethnic enclave (in tables A.51, A.58 and A.64) and average level of school in the origin country the year of birth of the father (in tables A.50, A.57 and A.63) and the year of moving to France (for trend in the country of origin) in tables A.49, A.56 and A.62.

5 A multi-generational model of immigrants human capital

Although the empirical evidence shown so far is easily relatable, it is not obvious what are the economic channels at play. Having a more formal framework helps highlighting mechanisms. In this section, I develop a model of human capital accumulation in the destination country. A characteristic of the model is that parents cannot fully transfer HC between labor markets of different countries but can transmit it to their children. To explain why rates of HC accumulation are different between the two generations, I focus on two alternative scenarios. In the first one, parents (who suffer from a loss of HC at migration) invest less into accumulating back HC for themselves than they invet into their children, but the technology governing the productivity of the two investments is the same. In the second scenario, the level of investments are the same but the technology is different: it is less productive for parents to invest in themselves than in their children.

To be easy to read, the evidence presented so far uses a binary definition of pre-migration status. Although the results hold when the definition is modified, it is never clear where the line

separating the sample in two should be drawn. For every definition used, it can always be argued that another would be better. When using a model, the discussion becomes more general and the empirical tests can be based on more common measures of human capital level (education level, years of experience in the labor market ...) and of investment (learning the language, helping children with homework ...).

5.1 Modelling framework

The main distinctive element of this model is that the economic assimilation of immigrants happens not only through the first but also the second generation. It keeps the general structure of Borjas (2015) but adds another generation. There are two periods. In the first one parents are adults. They arrive in the destination country and work. They cannot transfer the totality of their HC and lose a share $\delta \in (0, 1)$ of their raw level when they integrate the labor market of the destination country. In this period, children are young and not yet on the labor market.

In the first period, parents face two investment decisions: how much to invest into accumulating back HC and how much into developing the HC of their children. The costs of these investments is that they eat part of their HC in the first period. π is the share invested in adapting, i.e. *parents investment* and θ the share invested in developing the HC of children, i.e. *children investments*. Parents are benevolent, the destiny of their children matters for them as much as their own.

In the second period, both children and parents are on the labor market. Parents have accumulated extra HC since the first period, at rate g, meaning they enjoy $(1+g)$ of what they had in the first period. For simplicity, children start with the HC level of their parents, in the sense that there is full transmission[11]. Their HC between the two periods has grown at rate m.

I made two simplifying assumptions, (i) HC is unidimensional and (ii) there is no crowding out between π and θ. An alternative to (i) would have been to distinguish elements of HC that can be transmitted regardless of the socio-economic situation in the destination country from those that are dependent of it. For instance, educational aspirations or certain behavioral traits, such as self-control or patience could be in the former category. However, the data does not contain such detailed information and does not allow to dig in that direction. It is also not possible to assess the (potential) trade-off between investing for yourself and investing for your children as these informations do not appear together in the same sample. Since the purpose of the model is to illustrate the empirical evidence reported and motivate tests, it is important to keep the modelling framework in line with the data.

There are three important dimensions in the model. For each of them, the main modelling decision is to determine whether K or $(1-\delta)K$, the initial or the transported level of HC, is the relevant one. The three dimensions are: (i) the technology of HC production and in particular

[11]For ease of presentation, I also assume that parents value the second period as much as the first, i.e. have a discount factor of one.

which HC level determines the productivity of investments, (ii) the starting point over which HC is accumulated and (iii) from which measure of HC investment is eaten away.

I will go over the three points successively, detail the equations and provide some intuition. Following Ben-Porath (1967) and Borjas (2015) the rates of accumulation are non linear in the initial level of capital. The same investment implies higher accumulation for someone with a higher level of HC.

The technology of HC production For parents and children, it is driven by the initial level. If our fictional taxi driver wanted to take an exam to have his engineer background recognized, he would learn and work with the productivity of an engineer not that of a taxi driver. When he invests in his children, for instance by helping them with homework, he also has the productivity of an engineer.

The starting point from which HC is accumulated For parents, it is from $(1-\delta)K$. In the destination country, the father is a taxi driver and his situation will improve (or not) from that point. On the other hand, I assume that children start from the initial level of their family. The downgrading that happened in the labor market did not affect the situation inside the household.

Put together this gives the two following equations of investment's productivity, for *parents investment*:

$$g(1-\delta)K = (\pi K)^{\alpha} K^{\beta}$$

where g is the rate at which HC is accumulated from the starting point $(1-\delta)K$. The productivity for accumulation is determined by K. Following the same reasoning the equivalent equation for *children investment* is

$$mK = (\theta K)^{\xi} K^{\lambda}$$

In the model, α and ξ capture the transformation of investment into future HC while β and λ capture how much initial HC matters for accumulation.

From which HC level are investments taken away Since the investments are made by parents in the destination country in the period following their arrival, the available HC is $(1-\delta)K$. I assume an interior solution under the conditions that π, θ, α and ξ are inferior to one to avoid having zero or negative marketable capital in the first period.

The maximization program for parents in the first period is:

$$\max_{\pi,\theta} \left[(1-\delta)K - (\theta+\pi)(1-\delta)K\right] + \left[(1+g)(1-\delta)K + (1+m)K\right]$$

In the first period, the father enjoys the monetary equivalents of being a taxi driver $(1-\delta)K$ minus his investments $(\theta+\pi)(1-\delta)K$. Since only parents work in the first period, they do not benefit from the marketable HC of their children. In the second period, both have accumulated HC at their respective rates and from their respective starting points. The solutions are:

$$\pi^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\alpha}} \alpha^{\frac{1}{1-\alpha}} K^{\frac{\alpha+\beta-1}{1-\alpha}}$$

$$\theta^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\xi}} \xi^{\frac{1}{1-\xi}} K^{\frac{\xi+\lambda-1}{1-\xi}}$$

5.2 Two alternative scenarios

I take this result as a starting point to present two mutually exclusive scenarios. For ease of presentation, I focus on the case where K=1. The equations simplify to $\pi^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\alpha}} \alpha^{\frac{1}{1-\alpha}}$ and $\theta^* = \left(\frac{1}{1-\delta}\right)^{\frac{1}{1-\xi}} \xi^{\frac{1}{1-\xi}}$. The change in situation for the parents between the two periods is measured by

$$(1-\delta)(1+g^*) - (1-\delta) = g^*(1-\delta) \tag{3.3}$$

and the change between the parents and the children in the second period is measured by

$$(1+m^*) - (1+g^*)(1-\delta) = (m^*-g^*) + \delta(1+g^*) \tag{3.4}$$

The very low level of occupational mobility for the parents gives an indication that $g^*(1-\delta) \approx 0$. If I focus on the extreme case where $g^* = 0$, equation 3.4 shows the importance of including the loss of HC when measuring the mobility between parents and children. If δ is assumed to be zero, the ability of second generation to accumulate HC is overestimated. One crucial dimension of the dynamics that is not fully explained by the evidence presented so far is the term $(m^* - g^*)$. How different is the rate of accumulation of HC between parents and children and why?

One crucial element of the model is that the rate of HC accumulation for both parents and children can be rewritten as a function of the level of investments and a technological parameter. In particular, by plugging π^* and θ^* into the formulas for g and m, one obtains $\xi m^* = (1-\delta)\theta^*$ and $\alpha g^* = \pi^*$. This allows to focus on two alternative scenarios: the difference between the two rates can be driven by (i) a difference in level of investments or by (ii) a difference in the productivity of these investments. Is it (i) that the cab driver invests more in his children than in himself or (ii) that investing in his children pays off when investing in himself does not? In the first scenario, the investment levels are different but the technology is the same. In the second one, the investments levels are the same but the technology is different. In the framework of the model:

In scenario 1, $\xi = \alpha$ and $m^* > g^*$ can be rewritten as

$$m^* = \theta^*(1-\delta) > \pi^* = g^* \rightarrow \theta^* > \pi^* \tag{3.5}$$

So if the technology is the same than the difference is driven by differences in investments' quantity.

In scenario 2, the investment decisions are the same, $\theta^* = \pi^*$ and $m^* > g^*$ can be rewritten as

$$m^* = \frac{1-\delta}{\xi} > \frac{1}{\alpha} = g^* \rightarrow \alpha > \xi \tag{3.6}$$

Recall that if a number between 0 and 1 is elevated to a power itself between 0 and 1, a larger value of the exponent means a lower productivity, i.e. $\pi^\xi > \pi^\alpha$ if $\xi, \alpha, \pi \in (0,1)$ and $\alpha > \xi$. In the second scenario, investments are the same but productivity is higher for children investments. There are many combinations of differences in investment decisions and technology parameters that can also explain equation 3.4. Focusing on two extreme cases makes the channels more salient and motivate easy empirical tests.

6 Testing empirically the two scenarios

To compare the two scenarios, one needs information on the level of *parents* and *children investments* together with some measures of their productivity. To measure *parents investments*, I use information on whether (first generation) immigrants took French classes since their arrival. This information is available in *sample first*. To measure children investments, I switch to *sample second* and use its various measures of investments in time and money. To assess the productivity of these investments, I focus on their returns in the labor market or their effects on educational outcomes. I cannot strictly speaking compare the levels of investment, neither their returns, between parents and children as they are expressed in a different metric. I can however show evidence of investment level being high or low in absolute values and their returns being positive or null. To assess how productive they are, I can also compare, for certain investments, returns for natives and immigrants. Last but not least, I can also check that investment decisions are in line with the predictions of the model.

6.1 Scenario 1: Difference in investment levels

Descriptive evidence on investment levels Fluency in the destination country language has been shown in Chiswick (1991) to be associated with higher wages. Trying to improve one's language skills is thus a relevant investment (also mentioned in Borjas (2015)). TeO contains information on fluency at arrival in France and at the time of the interview. For respondents answering not being proficient in French, respectively 80% and 84% of male and female immigrants

with adult children[12], information is collected on whether they took French classes since their arrival. A large fraction made this investment since 33% of male immigrants and 46% of female immigrants report having followed classes in French. Those raw numbers are indicative that immigrants invest into adapting their human capital.

When it comes to children, I rely on the answers to the following questions, (i) were the children sent to a school outside their district (denoting that parents elaborate a schooling strategy), (ii) did parents often help them with their homework, (iii) did children have a place on their own to study and (iv) did parents pay for private classes outside school. These four variables capture key dimensions of parental investments in time, money and in elaborating a successful strategy for their children. Mean levels of investments for immigrants and natives are reported in table 10. The means are also expressed conditional on occupational level. Investments levels are lower for immigrants than natives but the raw levels are not strikingly different, especially when one compares $S_{P,2} = L$ parents, 17% vs 20% for schooling decisions or 66% vs 70% for having a room on their own to study. The only major difference is for helping with homework, native parents are twice more likely to help. For this investment immigrant parents probably face more difficulties than natives. Both sets of evidence taken together show that parents do invest in themselves and in their children.

[Table 10 here]

Evidence on who is investing more The model has one interesting prediction on how pre-migration background influences investment decisions. Parents who suffered from a larger downward shock should invest more in both themselves and their children, i.e. $\frac{\partial \pi^*}{\partial \delta} > 0$ and $\frac{\partial \theta^*}{\partial \delta} > 0$.

To test this prediction, I estimate the following equation:

$$y_i = \alpha + \sum_{j=1}^{4} \beta_j \text{educ}_i + \gamma X_i + \varepsilon_i \tag{3.7}$$

where y_i is a measure of investments, $educ_i$ are four dummy variables for level of education (below primary school, primary school, secondary school and higher education) and X_i are individual level controls. I estimate equation 3.7 on the subsample of immigrants who had a "low" status upon arrival, $S_{P,1} = L$. For a similar (low) socio-professional status, people with a higher level of education suffered from a stronger downward shock.

If the prediction is correct, the $\{\beta_j\}$ should not be equal. Results are reported in the first two columns of table 11. Since I focus on investments that should pay off in the labor market, I restrict the sample to male immigrants. Both specifications control for age and the one reported in the second column controls for the initial level in French[13]. There is a clear break between those with

[12] I focus on immigrants with adult children since this group produced the motivating evidence of section 3.

[13] I use a categorical variable coded in the survey and which takes four values.

secondary education and above. The coefficients for higher education are not precisely estimated since few immigrants (only 48) with that educational background are part of the $S_{P,1} = L$. I can also reject the null hypothesis of equal coefficients at standard levels.

[Table 11 here]

To test this hypothesis on *sample second*, I estimate equation 3.7 on the subsample of immigrants whose fathers have a "low" occupation, $S_{P,2} = L$. The results are reported in table 11, each column refers to a type of investment. There is clear evidence that for a similar low status, those who fell from higher (because they had a higher level of education) invested more in their children. I can reject the null of the equality of coefficients at the 1% level for all forms of investment but elaborating a schooling strategy (at the 10% level). Compared to *parents investment*, there is no break below/above secondary education but a more gradual increase. These results show the importance of incorporating pre-migration information to understand the process of human capital accumulation for immigrants. It is also reassuring to see that the predictions of the model are in line with the data.

6.2 Scenario 2: Difference in investments productivity

In this subsection, I test the implications of the second scenario. Under it, the problem for first generation immigrants is not so much that they do not invest in themselves but that it is less productive than it is to invest in their children.

Investments do not pay off for the first generation. I look at the productivity of two investments in human capital for immigrants: an indirect one, the returns to experience in the host country and a direct one, language courses. To test the former, I estimate

$$\ln(y_i) = \alpha + \beta_1 \text{exp}_{\text{FR}} + \beta_2 \text{exp}^2_{\text{FR}} + \beta_3 \text{exp}_{\text{before}} + \beta_4 \text{exp}^2_{\text{before}} + \gamma X_i + \varepsilon_i \tag{3.8}$$

from Friedberg (2000) where exp_{FR} refers to the number of years in France, $\text{exp}_{\text{before}}$ the difference between the age at migration and 18 and y the wage. I estimate this regression on males and report the results in the upper part of table 12. I include country of origin fixed effects in column 2 and 5 and control for education level in all specifications. To give an idea of how small/large the returns are, I report in column 3and 6 the rates of return to experience[14] for native males. In the baseline specification, the returns are negative for immigrants. Although this result is surprising, it is important to note that it is based on a sample of immigrants who have adult children. For instance, when re-estimated on all immigrants (and not only those who have adult children), the coefficient is positive, a little above 1% which is very close to the initial result by Friedberg (2000).

[14] Measured by age-18.

When I include additional controls such as year of arrival (see tables A.63), the coefficient becomes insignificant. What always hold (and what is important here) is that returns are much lower than for natives (column 3).

To check that this result is not specific to one measure of investment in accumulating back human capital, I estimate equation 3.9 on male immigrants and report the results in the lower part of table 12.

$$\ln(y_i) = \alpha + \beta_1 \text{course} + \gamma X_i + \varepsilon_i \tag{3.9}$$

Course is a dummy taking value one of the immigrant took language classes. I use the same controls as for estimation of equation 3.8. The effect is not statistically different from zero. This does not mean that being fluent is not associated with higher wages but that long term effects of having taken French classes are not quantitatively important. These results differ from the high returns found in Cohen-Goldner and Eckstein (2008) on estimations 5 years after arrival in Israel. In my setting, I estimate longer term returns. This result does not change when I estimate equation 3.9 on all immigrants (and not only those who have children), or by including year of arrival fixed effects (see table A.63).

[Table 12 here]

Investments for second generation immigrants are more productive. For each potential outcome used to show resurgence and every measure of investment available in *sample second*, I regress the outcome on the investment interacted with a dummy for immigrants and natives (and a series of individual controls, namely age, gender, country of origin and socio-economic status of the father). This gives the following equation to estimate:

$$y_i = \alpha + \sum_{j=0}^{1}\sum_{l=0}^{1} \beta_{j,l} 1\left\{\text{Native}_i = j, \text{Investment}_i = l\right\} + \gamma X_i + \varepsilon_i \tag{3.10}$$

This allows to test whether the returns to investment are positive but also if they are lower, larger or similar than for natives. Results are reported in table 13. Each panel focuses on a type of investment, each row on an outcome.

[Table 13 here]

Returns to parental investment for immigrants are particularly high for educational outcomes. They are either indistinguishable or if anything higher than for natives. The last rows of each panel report the p-value of the F-test associated with the null of equality between the two coefficients. It fails to reject the null (at the 5% level) in 11 out of the 20 regressions. In the nine remaining cases, coefficients for immigrants are larger than for natives. Returns are sometimes negative for natives which probably captures reverse causality. This table indicates that immigrant parents do not face

the same problems developing the human capital of their children than they have adapting their own.

7 Conclusion

In this paper, I include information on the socio-economic status of the family pre-migration and see how it affects the understanding of the economic and educational achievements of first and second generation immigrants. Using unique French survey data, I reconstitute family histories over three generations, partially in the origin country, partially in France.

I find that there is a dichotomy between what happens in the labor market and in the household after migration. While first generation immigrants lose part of their HC in the labor market and have difficulties accumulating back HC in the destination country; their children still benefit from any pre-migration advantage. This extra dimension is then a key element in the understanding of the situation of second generation immigrants.

One of the limitation of this paper is that it does not allow to distinguish between the different elements of HC. In particular, no information is available on the transmission of behaviors and aspiration. Future research could push in that direction.

Bibliography

Algan, Y., Dustmann, C., Glitz, A., and Manning, A. (2010). The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. *Economic Journal*, 120(542):F4–F30.

Aydemir, A., Chen, W.-H., and Corak, M. (2009). Intergenerational Earnings Mobility among the Children of Canadian Immigrants. *The Review of Economics and Statistics*, 91(2):377–397.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*, 104(C):184–198.

Becker, G. S. and Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4(3):S1–39.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4):352–365.

Blau, F. D. and Kahn, L. M. (2015). Substitution between Individual and Source Country Characteristics: Social Capital, Culture, and US Labor Market Outcomes among Immigrant Women. *Journal of Human Capital*, 9(4):439 – 482.

Borjas, G. J. (1993). The Intergenerational Mobility of Immigrants. *Journal of Labor Economics*, 11(1):113–35.

Borjas, G. J. (1995). Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s? *Journal of Labor Economics*, 13(2):201–245.

Borjas, G. J. (2015). The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again. *Journal of Human Capital*, 9(5):483–517.

Braun, S. T. and Stuhler, J. (2018). The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany. *Economic Journal*, 128(609):576–611.

Caponi, V. (2011). Intergenerational Transmission Of Abilities And Self?Selection Of Mexican Immigrants. *International Economic Review*, 52(2):523–547.

Card, D. (2005). Is the New Immigration Really so Bad? *Economic Journal*, 115(507):300–323.

Chan, T. W. and Bolivier, V. (2013). The Grandparents Effect in Social Mobility: Evidence from British Cohort Studies. *American Sociological Review*, 92(1):662–679.

Chiquiar, D. and Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, 113(2):239–281.

Chiswick, B. R. (1991). Speaking, Reading, and Earnings among Low-Skilled Immigrants. *Journal of Labor Economics*, 9(2):149–170.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Number 10181 in Economics Books. Princeton University Press.

Cohen-Goldner, S. and Eckstein, Z. (2008). Labor Mobility Of Immigrants: Training, Experience, Language, And Opportunities. *International Economic Review*, 49(3):837–872.

Eckstein, Z. and Weiss, Y. (2004). On The Wage Growth of Immigrants: Israel, 1990-2000. *Journal of the European Economic Association*, 2(4):665–695.

Fernandez, R. (2007). Women, Work and Culture. *Journal of the European Economic Association*, 5(2-3):305–332.

Fogli, A. and Fernandez, R. (2006). Fertility: The Role of Culture and Family Experience. *Journal of the European Economic Association*, 4(2-3):552–561.

Friedberg, R. M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, 18(2):221–251.

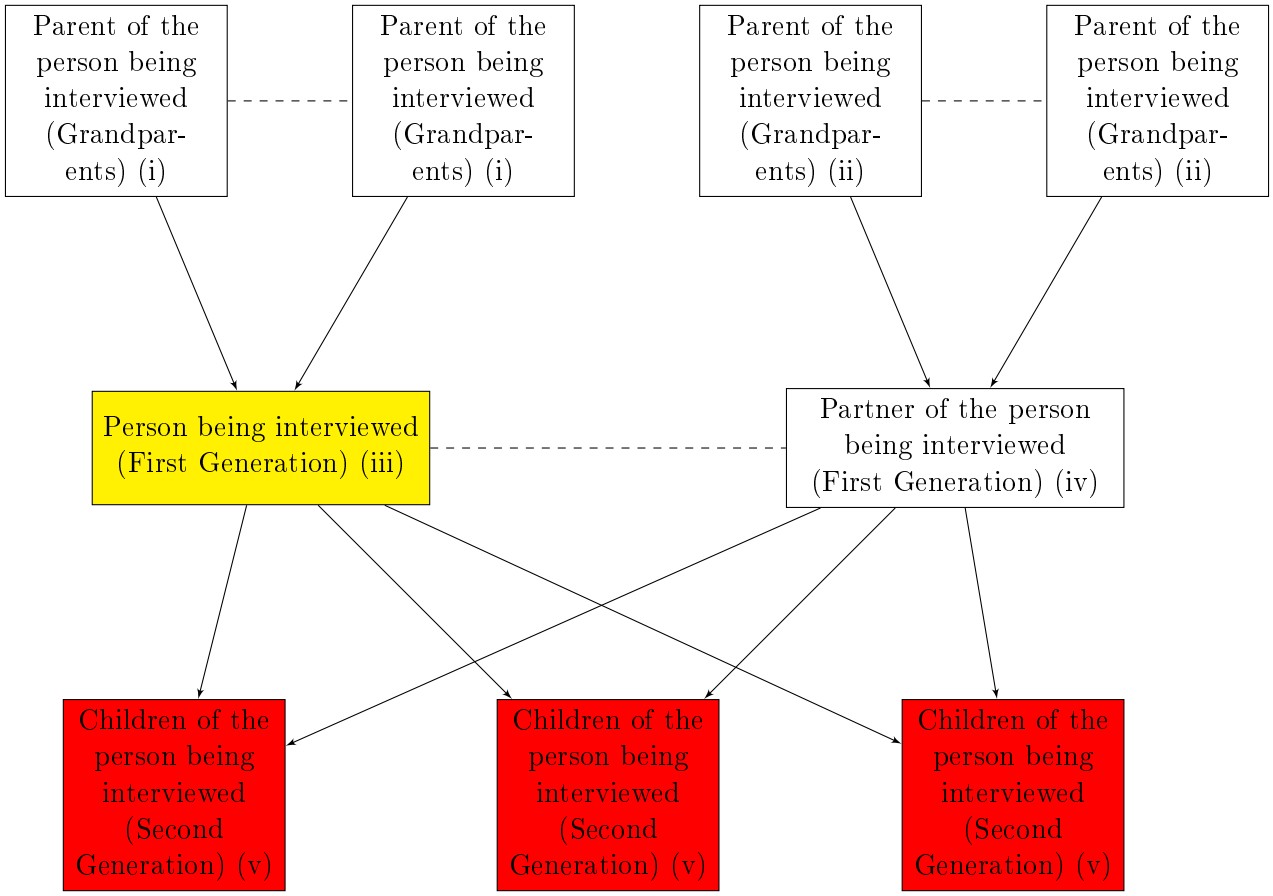
Gang, I. N. and Zimmermann, K. F. (2000). Is Child like Parent? Educational Attainment and Ethnic Origin. *Journal of Human Resources*, 35(3):550–569.

Ichou, M. (2014). Who They Were There: Immigrants' Educational Selectivity and Their Children's Educational Attainment. *European Sociological Review*, 30(6):750 – 765.

Kaestner, R. and Malamud, O. (2014). Self-Selection and International Migration: New Evidence from Mexico. *The Review of Economics and Statistics*, 96(1):78–91.

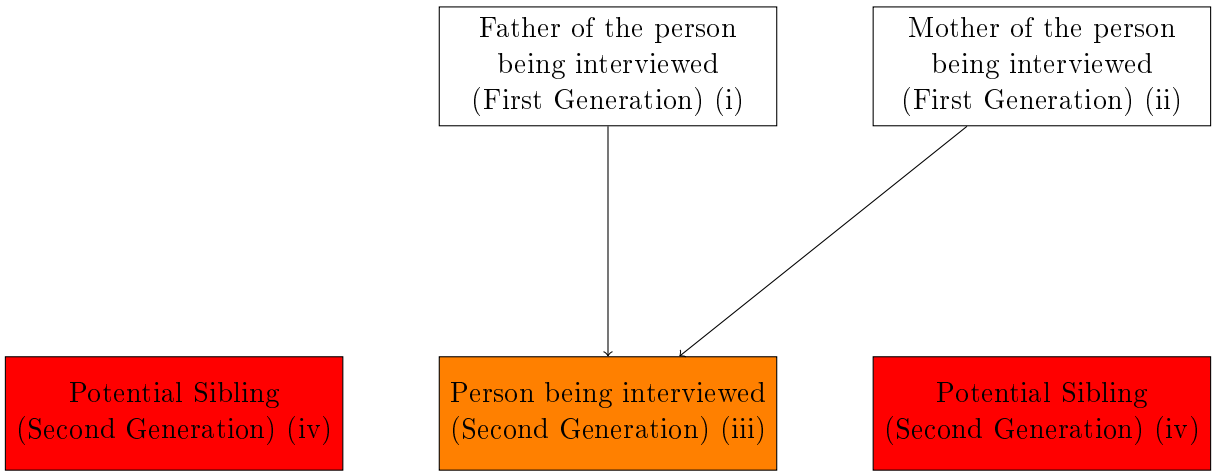
Mare, R. (2011). A Multigenerational View of Inequality. *Demography*, 48(1):1–23.

Figure 1: Who is who? (sample first)



In Figure 1 The person being interviewed in TeO is in yellow, the population of interest in red.

Figure 2: Who is who? (sample second)



In Figure 2 The person being interviewed in TeO is in yellow, the population of interest in red.

Since the person interviewed is also from the population of interest, it is orange.

Table 1: What information is available for whom?

		Part of the Survey		
		Personal History	Current Situation	Children
Sampled as	first	grandparents 0G parents pre mig 1st G	parents France 1st G	children France 2nd G
	second	parents pre mig 1st G	parents France 1st G children France 2nd G	grandchildren France 3rd G

		Part of the Survey		
		Personal History	Current Situation	Children
Sampled as	first	occupation, education occupation, education	occupation, wage, transition on L market	education employment
	second	education	occupation education, occupation, wage	education employment

Note: In each cell is described the population targeted by looking at the sample in the survey (rows) and the questions asked to him/her (columns). In the upper table, I report the population, in the lower one the type of information collected.

Table 2: Descriptive statistics - First Generation Immigrants - *Sample first*

Gender		
Women	1266	59%
Men	873	41%
Year of Birth		
25th percentile	1952	
Median	1956	
75th percentile	1961	
Education Level		
< Primary School	661	31%
Primary School	608	28%
Secondary Scool	270	13%
Higher Education	390	18%
Not Available	210	10%
Occupation Pre-Mig		
Self Employed Agricultural	80	4%
Self Employed Non-Agricultural	121	6%
High Managerial	114	5%
Supervisory Occupations	174	8%
Lower Services	259	12%
Lower Technical	358	17%
Not Available	1033	48%
Education level father		
< Primary School	1307	61%
Primary School	315	15%
Secondary Education	121	6%
Higher Education	199	9%
Not available	197	9%
Occupation Father		
Self Employed Agricultural	427	20%
Self Employed Non-Agricultural	388	18%
High Managerial	188	9%
Supervisory Occupations	149	7%
Lower Services	217	10%
Lower Technical	666	31%
Not Available	104	5%

Note : The Observations are parents of second generation immigrants followed in sample first. I report the gender of the parent being interviewed, year of birth, education level and occupation before migration. I also report the occupation and education level of their father, i.e. the grandfather of second generation immigrants. The entry 1307 should be read as follows; among first generation immigrants, 1307 had a father who completed less than primary school.

Table 3: Descriptive statistics - Status in the country of origin - *Sample first* - *Parents definition*

Country of origin	$S_{P,0} = L$	$S_{P,0} = H$	Nb of Observations
Algeria	57	43	61
Germany	24	76	38
Central America	39	61	18
North America	8	92	13
South America	20	80	30
Africa (Other)	38	62	40
Europe (Other)	42	58	53
Belgium	24	76	51
Cambodia	72	28	36
Cameroun	75	25	24
Congo B	50	50	16
Ivory Cost	45	55	11
Spain	71	29	14
Italy	70	30	27
Laos	54	46	35
Mali	81	19	21
Marocco	79	21	82
Middle East	6	94	33
Poland	32	68	19
Portugal	95	5	166
RDC	19	81	36
Senegal	76	24	29
Tunisia	78	22	36
Turkey	82	18	80
UK	16	84	51
Vietnam	43	57	58
Asia (Other)	23	77	52

Note : The Observations are parents of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. The table only reports countries with more than 15 immigrant parents.

Table 4: Descriptive statistics - Status at the junction of the countries of origin and of immigration

	Mean			Number of Obs		
	$S_{P,0}$=H $S_{P,2}$=H	$S_{P,0}$=H $S_{P,2}$=L	$S_{P,0}$=L $S_{P,2}$=L			
	(1)	(2)	(3)	(1)	(2)	(3)
		Europe				
Year of arrival	1989	1989	1980	137	83	230
Male	0.36	0.35	0.50	137	83	230
Fluency in French	0.40	0.32	0.07	116	77	222
Secondary Education	0.95	0.89	0.01	137	83	227
Higher Education	0.65	0.30	0.00	137	83	227
Occupation Grandfather	Retailers and Related	Skilled Worker	Farmers	137	83	230
Occupation Fathers	Managerial Occupations - Business	Skilled Industrial Workers	Skilled Industrial Workers	50	29	116
Occupation Mothers	Managerial Occupations	Workers - direct services	Workers - direct services	87	54	114
		Asia				
Year of arrival	1986	1984	1983	38	75	142
Male	0.42	0.51	0.61	38	75	142
Fluency in French	0.22	0.15	0.06	36	72	139
Secondary Education	0.95	0.89	0.00	38	75	134
Higher Education	0.61	0.27	0.00	38	75	134
Occupation Grandfather	Retailers and Related	Farmers	Farmers	38	75	142
Occupation Fathers	Managerial Occupations - Business	Skilled Industrial Workers	Skilled Industrial Workers	16	38	87
Occupation Mothers	No activity	Workers - direct services	Unskilled Industrial Workers	22	37	55
		Maghreb				
Year of arrival	1982	1982	1979	14	37	121
Male	0.71	0.59	0.68	14	37	121
Fluency in French	0.85	0.78	0.42	13	32	116
Secondary Education	0.79	0.70	0.02	14	37	97
Higher Education	0.64	0.27	0.00	14	37	97
Occupation Grandfather	Intermediate Administrative - Gov	Retailers and Related	Farmers	14	37	121
Occupation Fathers	Intermediate Administrative - Business	Skilled Industrial Workers	Skilled Industrial Workers	10	22	82
Occupation Mothers	No activity	Lower Administrative - Gov	Workers - direct services	4	15	39
		Africa				
Year of arrival	1991	1989	1987	34	67	104
Male	0.59	0.48	0.38	34	67	105
Fluency in French	0.93	0.85	0.62	30	60	93
Secondary Education	0.91	0.82	0.07	34	67	92
Higher Education	0.53	0.37	0.04	34	67	92
Occupation Grandfather	Retailers and Related	Farmers	Farmers	34	67	105
Occupation Fathers	Intermediate Administrative - Gov	Skilled Industrial Workers	Skilled Industrial Workers	20	32	40
Occupation Mothers	No activity	Workers - direct services	Workers - direct services	14	35	65

Note : The Observations are parents of second generation immigrants followed in sample first. High and Low in period 2 refer to the occupation of the father. High and Low in period 0 refer to the parent definition used for sample first. For different geographical areas, I report the mean year of arrival, the proportion of men, the level of spoken French at the time of arrival, the proportion of immigrants having completed secondary or higher education as well as the mode(s) of occupation of grandfathers in the country of origin and mode(s) of occupation of the parents in France.

Table 5: Transition matrix - Before migration to first job - *Parents definition*

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	340	10	350	97.14	2.86	58.53
$S_{P,0} = H$	157	91	248	63.31	36.69	41.47
Total	497	101	598	83.11	16.89	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father.

Table 6: Transition matrix - From arrival to current job - *Parents definition*

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	446	51	497	89.74	10.26	83.11
$S_{P,1} = H$	14	87	101	13.86	86.14	16.89
Total	460	138	598	76.92	23.08	100
Natives	609	419	1028	59.24	40.76	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18.

Table 7: Transition matrix - From arrival to current job - *Parents definition* - Previously High Status

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	119	38	157	75.80	24.20	63.31
$S_{P,1} = H$	12	79	91	13.19	86.81	36.69
Total	131	117	248	52.82	47.18	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The sample is limited to fathers with a high status in the origin country.

Table 8: Resurgence - *Sample First - Parents Definition*

Obtaining the Baccalaureat

$S_{P,0} = H$	0.133***	0.129***	0.157***		
	(0.033)	(0.044)	(0.036)		
$S_{P,1} = H$		0.089*			
		(0.056)			
$S_{P,2} = H$	0.136***				
	(0.033)				
Houhesold $			0.002**		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				-0.060	
				(0.098)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.111***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.245***	
				(0.069)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.003
					(0.065)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.099***
					(0.035)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.295***
					(0.038)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 3.1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 3.2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender and dummies for country of origin. Standard errors are clustered at the household level.

Table 9: Resurgence - *Sample Second*

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.08***	0.17***	0.26***	0.14	0.21***
	(0.02)	(0.03)	(0.05)	(0.11)	(0.08)
$S_{P,2} = H$	0.03*	0.13***	0.13***	0.16**	0.20***
	(0.02)	(0.03)	(0.05)	(0.07)	(0.08)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.53	0.32	7.40	0.37
N	3339	3339	1933	667	859
N High	397	397	143	46	63
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.04*	0.17***	0.15***	0.24***	0.26***
	(0.03)	(0.04)	(0.06)	(0.07)	(0.09)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.08***	0.19***	0.29***	0.27**	0.29***
	(0.02)	(0.04)	(0.06)	(0.15)	(0.10)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.10***	0.28***	0.35***	0.17**	0.34***
	(0.01)	(0.03)	(0.06)	(0.10)	(0.10)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.53	0.32	7.40	0.37
N	3339	3339	1933	667	859
N High	397	397	143	46	63

Note : The upper part of the table summarizes the results of the estimation of equation 3.1. The lower part of the table reports the results of estimating equation 3.2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender and dummies for country of origin. Standard errors are robust to general forms of heteroskedasticity.

Table 10: Descriptive Statistics - Children investment

	School	Homework	Room	Private Classes
	Among Immigrants			
$S_{P,2} = L \,\&\, S_{P,2} = H$	0.176	0.331	0.675	0.139
$S_{P,2} = L$	0.168	0.299	0.662	0.118
	Among Natives			
$S_{P,2} = L \,\&\, S_{P,2} = H$	0.234	0.659	0.758	0.228
$S_{P,2} = L$	0.201	0.615	0.701	0.168

Note: The table shows mean levels of parental investments. The four investments are sending children to a school outside the neighborhood "School", helping with homework "Homework", providing a school where children can study alone "Room" and pay for private classes "Private Classes". Respondents are adult second generation immigrants from *sample second* and natives.

Table 11: Predictions of the model - Parents and Children investment

	Language		School	Homework	Room	Private Classes
	$S_{P,1} = L$		$S_{P,2} = L$			
Primary School	-0.023	-0.026	0.031**	0.220***	0.082***	0.048***
	(0.046)	(0.046)	(0.017)	(0.019)	(0.021)	(0.014)
Secondary School	0.155**	0.176***	0.065**	0.315***	0.104***	0.103***
	(0.070)	(0.069)	(0.035)	(0.040)	(0.043)	(0.030)
Higher Education	0.097	0.120*	-0.017	0.443***	0.238***	0.123***
	(0.078)	(0.077)	(0.048)	(0.055)	(0.059)	(0.041)
Control Age	Yes	Yes	Yes	Yes	Yes	Yes
Control Gender	No	No	Yes	Yes	Yes	Yes
Country FE	No	No	Yes	Yes	Yes	Yes
Initial level FE	No	Yes	No	No	No	No
Nb observations	518	518	3,048	2,998	3,065	3,065
p-value H_0 equality	0.049	0.014	0.087	0.000	0.000	0.000
Mean	0.32	0.32	0.168	0.299	0.662	0.118

Note: This table shows the results of the estimation of equation 3.7. The left part of the table estimates it on the sample of "low" occupation, i.e. $S_{P,1} = L$ fathers of second generation immigrants in *sample first*. The outcome variable is having taken French classes. The first column controls for age and the second also include dummy variables for the level of French at the time of arrival (categorical variable provided by TeO and taking four values). The right part of the table reports the results on second generation immigrants whose fathers occupation in France is "low", $S_{P,2} = L$. There are four outcome variables, sending children to a school outside the neighborhood "School", helping with homework "Homework", providing a school where children can study alone "Room" and pay for private classes "Private Classes". All specifications control for age and gender of the second generation immigrant and the country of origin of their father. The row labelled "p-value H_0" are the p-value of the test of the equality of coefficients between all education level.

Table 12: Returns to experience and Language classes

	Immigrants with children			All Immigrants		
	Immigrants		Natives	Immigrants		Natives
Experience	-0.031**	-0.025**	0.121***	0.011**	0.013**	0.045***
	(0.015)	(0.014)	(0.050)	(0.004)	(0.004)	(0.004)
Experience sq	0.001***	0.000***	-0.002**	-0.000***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
Nb observations	506	506	246	1747	1747	1213
Control Education	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes		No	Yes	
Language classes	0.023	0.038		-0.025	-0.002	
	(0.051)	(0.052)		(0.029)	(0.029)	
Control Education	Yes	Yes		Yes	Yes	
Control Age	Yes	Yes		Yes	Yes	
Country FE	No	Yes		No	Yes	
Nb observations	356	356		1158	1158	
N High *Parents definition*	87	87		360	360	
N High *Gdparents definition*	154	154		568	568	

Note: The first two columns of the upper part of the table report the results of the regression of equation 8. The coefficients reported refer to experience in France (calculated as the number of years since arrival). The sample comprises fathers of second generation immigrants. Controls include dummies for level of education. Certain specifications include dummies for country of origin. The third column estimates a similar specification for natives with adult children where potential experience is calculated as age minus 18. The last three columns of the upper part of the table replicates these results on a sample of all male immigrants and natives. The first two columns of the lower part of the table reports the estimation of equation 9 on a sample of fathers of second generation immigrants who were not fluent in French at the time of arrival. All specifications include dummies for education level and birth year of the father; some include country of origin fixed effects. The last two rows specifiy the number of $S_{P,0} = H$ in the sample according to both the parents and grandparents definition. The last three columns of the lower part of the table replicates these results on a sample of all male immigrants.

Table 13: Returns on Children investment

	Help for homework				
	D-O	BAC	H-E	Occupation	ln Wage
Immigrants	0.016*	0.026*	0.071***	0.042	-0.002
	(0.011)	(0.018)	(0.024)	(0.038)	(0.034)
Natives	-0.036**	0.014	-0.050**	0.007	-0.104***
	(0.020)	(0.028)	(0.028)	(0.038)	(0.020)
Nb observations	7,193	7,193	5,702	4,079	3,243
Nb of Immigrants	3,550	3,550	2,059	922	729
p-value H_0 equality	0.010	0.673	0.000	0.437	0.002
	Private Classes				
Immigrants	0.041***	0.083***	0.126***	0.080*	0.134***
	(0.013)	(0.023)	(0.032)	(0.054)	(0.050)
Natives	0.006	0.111***	0.017	0.085**	-0.006
	(0.020)	(0.030)	(0.032)	(0.040)	(0.028)
Nb observations	7,276	7,276	5,758	4,102	3,264
Nb of Immigrants	3,618	3,618	2,100	937	741
p-value H_0 equality	0.103	0.416	0.009	0.936	0.009
	Room alone				
Immigrants	0.055***	0.061***	0.054***	-0.032	-0.008
	(0.013)	(0.017)	(0.020)	(0.033)	(0.029)
Natives	0.008	0.079***	-0.015	0.014	-0.082***
	(0.022)	(0.029)	(0.029)	(0.041)	(0.026)
Nb observations	7,276	7,276	5,758	4,102	3,264
Nb of Immigrants	3,618	3,618	2,100	937	741
p-value H_0 equality	0.018	0.517	0.016	0.230	0.000
	Schooling Strategy				
Immigrants	0.031***	0.056***	0.089***	0.014	-0.001
	(0.013)	(0.021)	(0.029)	(0.044)	(0.040)
Natives	0.002	0.052**	-0.058**	0.058*	-0.074***
	(0.021)	(0.031)	(0.032)	(0.041)	(0.027)
Nb observations	7,161	7,161	5,652	4,019	3,195
Nb of Immigrants	3,587	3,587	2,078	925	730
p-value H_0 equality	0.201	0.915	0.000	0.430	0.088

Note: The table presents the results of equation 3.10. The outcomes are not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation". Estimation is performed on a sample of second generation immigrants and natives. The rows refered to as Immigrant and Natives are the coefficients associated with investing and being an immigrant or a native. The four investments are sending children to a school outside the neighborhood "School", helping with homework "Homework", providing a school where children can study alone "Room" and pay for private classes "Private Classes". The baseline category is being a non investing immigrant. Regressions control for age, gender, country of origin and socio-economic status of the father. In each subtable the row p-value H_0 tests the hypothesis of equality between the two coefficients reported.

8 Data appendix

The purpose of this data appendix is to:

1. Represent graphically to whom in the family the information used in both samples refer to.
2. Explain in more details how the definitions of "high" and "low" statuses are built and which restrictions are put for observations to be included or not in the samples.
3. Detail which of these restrictions are due to data availability and justify the other restrictions that are not due to data availability.
4. Answer the FAQ about the data.

Figure A1: Who is who? (sample first)

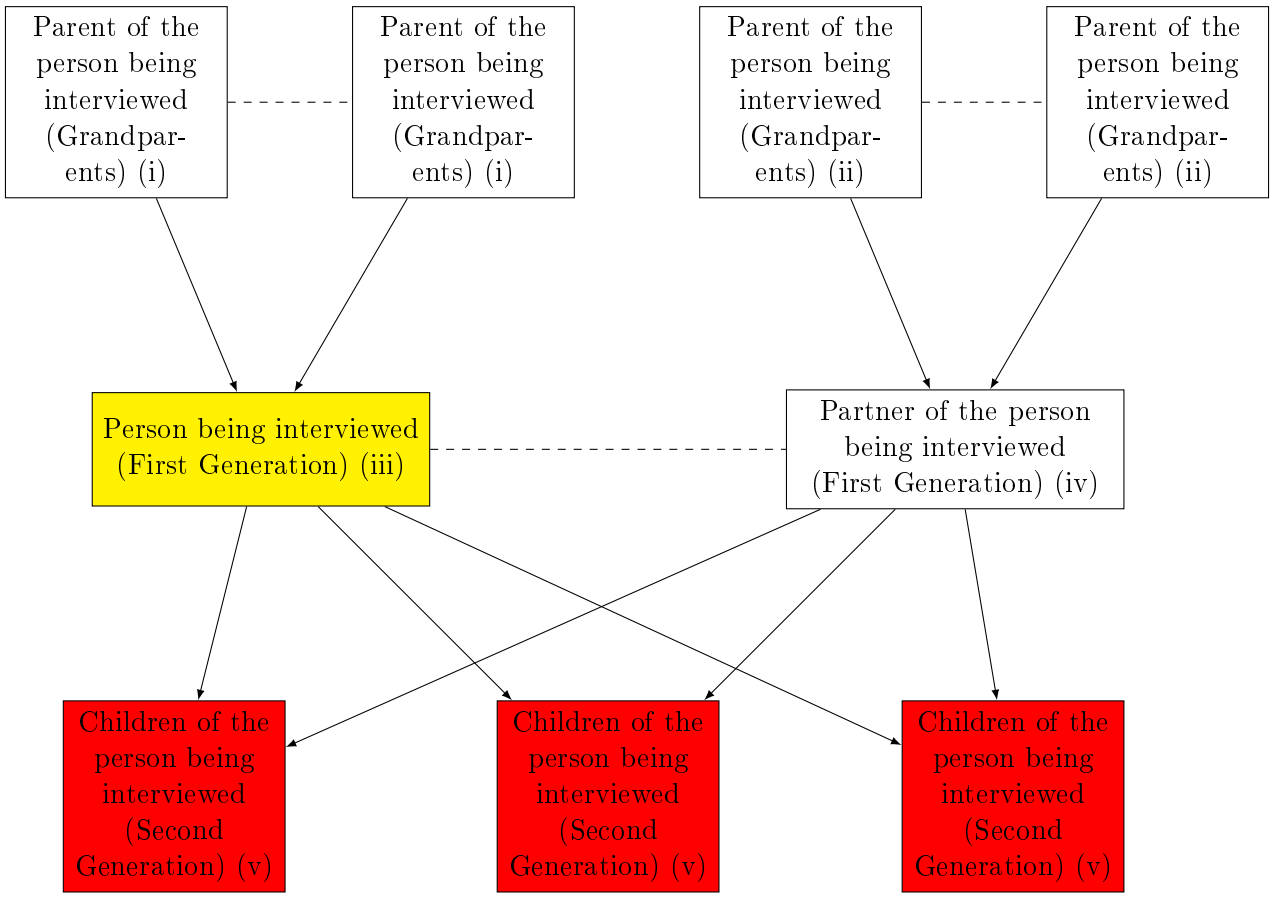


Figure A1: The person being interviewed in TeO is in yellow, the population of interest in red.

Subject	Criteria	Justification

To be part of the sample	(v) must have both parents, i.e. (iii) and (iv), immigrants	The reason why I choose to look at families with both parents immigrants (as opposed to those with one or two parents) is to make sure that I capture the influence of the pre-migration SES. By focusing on children of mixed couples (one immigrant and one native), the influence of the pre-migration status is mixed with the influence of the socio-economic background of the parent who is a native. It can then be that the relevant background for the children is the one of the native parent.
	(v) must have arrived before age 10	To make sure that they have been socialized in the environment characterized by $S_{p,2}$.
	(i) must have arrived after age 60 or not have emigrated to France	Since TeO is a cross section, the variable indicating the occupation of the grandfather in 2008 could be his occupation in France if he had moved. To prevent this case, I keep in the sample families whose grandfather never moved to France or moved after age 60. A caveat is that I can only restrict the sample on the basis of the location (or time of migration) of one of the grandfathers (the father of the parent of the person being interviewed)
Status in France	The National Statistical Agency (INSEE) has a one digit nomenclature of professional occupation with six main categories: Self Employed Agricultural, Self Employed Non Agricultural, High Managerial, Supervisory Occupations, Lower Services and Lower Technical. Are considered "high" status, the individuals whose occupation is classified as High Managerial or Supervisory Occupations, the rest is classified as "low" status.	
	Based on the occupation of the father (iii) if he is a man	
	Based on the occupation of the partner of the person being interviewed (iv) if (iii) is a woman provided (iii) reports partnership starting before the birth (v). I do not have information on the occupation of (iv) when (iii) arrived in France.	
Status before migration	Parents Definition occupation of (iii) or education level of (iii)	
	Grandparents Definition occupation of (i) (or (ii)), or education level of (i)	
Outcomes	Baccalauréat, being 18 or above	18 is a sensitive age to have finished high school.
	Employment, being 25 or above, being male	25 is a sensitive age to have finished university. I restrict the sample to males to account for the issues related to female participation.

Figure A2: Who is who? (sample second)

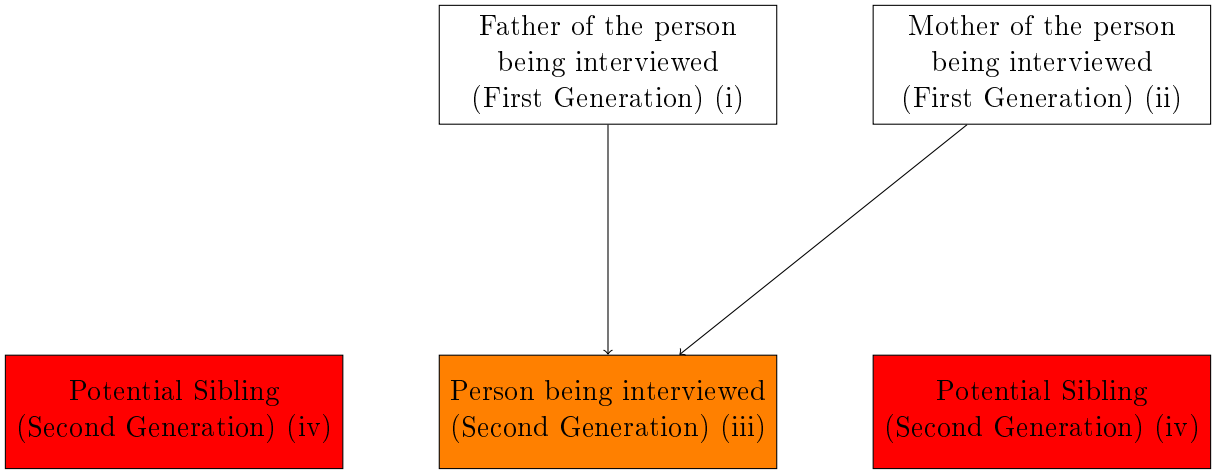


Figure A2: The person being interviewed in TeO is in yellow, the population of interest in red. Since the person interviewed is also from the population of interest, it is orange.

Category of data	Criteria	Justification
To be part of the sample	(iii) must have both parents	The reason why I choose to look at families with both parents immigrants (as opposed to those with one or two parents) is to make sure that I capture the influence of the pre-migration SES. By focusing on children of mixed couples (one immigrant and one native), the influence of the pre-migration status is mixed with the influence of the socio-economic background of the parent who is a native. It can then be that the relevant background for the children is the one of the native parent.
	(i) and (ii) have arrived after age 18	To make sure that parents received (or not) their primary education in their country of origin, I restrict the sample to parents who arrived after age 18
Status in France	The National Statistical Agency (INSEE) has a one digit nomenclature of professional occupation with six main categories: Self Employed Agricultural, Self Employed Non Agricultural, High Managerial, Supervisory Occupations, Lower Services and Lower Technical. Are considered "high" status, the individuals whose occupation is classified as High Managerial or Supervisory Occupations, the rest is classified as "low" status.	
Status before	(i) finished secondary education	
migration	Dropout and Baccalauréat, being 18 or above	18 is a sensitive age to have finished high school.
Outcomes	Employment, Wage, White Collar being 25 or above, being male	25 is a sensitive age to work, it coincides with the age set for having finished university ensuring no composition effects in our comparisons. I restrict the sample to males to account for the issues related to female participation.
	Higher Education, being 25 or above	25 is a sensitive age to have finished university.

Information on the additional survey	I create several dummy variables if the child has been in a school different than the one of their neighborhood where he was supposed to go, if children report that either their father or their mother helped them sometimes or often with their homework, if children had a room of their own to study and if parents paid for private classes to their children	

9 Descriptive Statistics on the different subsamples

Table A1: Descriptive statistics - Status in the country of origin - *Sample first* - *Grandarents definition*

Country of origin	$S_{P,0} = L$	$S_{P,0} = H$	Nb of Observations
Algeria	46	54	235
Germany	19	81	91
Central America	35	65	65
North America	9	91	46
South America	22	78	65
Africa (Other)	34	66	124
Europe (Other)	25	75	142
Belgium	17	83	106
Benin	47	53	17
Bulgaria	22	78	18
Burkina Fasso	60	40	10
Cambodia	58	42	64
Cameroun	54	46	65
Congo B	35	65	60
Ivory Cost	58	42	59
Spain	43	57	42
Guinee	59	41	27
Guinee Bissao	80	20	10
Italy	46	54	63
Laos	58	42	45
Mali	74	26	61
Marocco	57	43	250
Mauritania	50	50	16
Middle East	16	84	68
Netherlands	11	89	35
Poland	32	68	57
Portugal	91	9	261
RDC	21	79	81
Centrafrica	55	45	11
Romania	17	83	46
Senegal	61	39	103
Switzerland	25	75	28
Togo	74	26	19
Tunisia	61	39	105
Turkey	62	38	296
UK	17	83	127
Vietnam	38	62	120
Asia (Other)	27	73	200

Note : The Observations are parents of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first.

Table A2: Descriptive statistics - Status in the country of origin - *Sample first* - *Parents definition*

Country of origin	$S_{P,0} = L$	$S_{P,0} = H$	Nb of Observations
Algeria	75	25	168
Germany	20	80	40
Central America	55	45	29
North America	6	94	16
South America	24	76	33
Africa (Other)	30	70	63
Europe (Other)	39	61	70
Belgium	26	74	53
Cambodia	71	29	73
Cameroun	38	62	37
Congo B	38	62	32
Ivory Cost	58	42	26
Spain	64	36	36
Italy	68	32	44
Laos	64	36	75
Mali	71	29	45
Marocco	80	20	231
Middle East	17	83	48
Poland	43	57	23
Portugal	81	19	237
RDC	37	63	51
Senegal	77	23	71
Tunisia	76	24	92
Turkey	73	27	177
UK	17	83	54
Vietnam	41	59	95
Asia (Other)	25	75	73

Note : The Observations are first generation immigrants (with or without children) followed in sample first. High and Low in period 0 refer to the parents definition used for sample first.

Table A3: Descriptive statistics - Status in the country of origin - *Sample first* - *Grandparents definition*

Country of origin	$S_{P,0} = L$	$S_{P,0} = H$	Nb of Observations
Algeria	59	41	515
Germany	13	87	107
Central America	37	63	101
North America	2	98	56
South America	16	84	92
Africa (Other)	40	60	212
Europe (Other)	24	76	200
Belgium	21	79	130
Benin	30	70	33
Bulgaria	5	95	22
Burkina Fasso	45	55	20
Cambodia	54	46	146
Cameroun	29	71	122
Congo B	36	64	126
Ivory Cost	43	57	125
Spain	41	59	79
Guinee	60	40	60
Guinee Bissao	79	21	19
Italy	47	53	92
Laos	58	42	126
Mali	69	31	147
Marocco	65	35	631
Mauritania	70	30	33
Middle East	15	85	123
Netherlands	8	92	38
Poland	25	75	76
Portugal	76	24	394
RDC	29	71	133
Centrafrica	22	78	27
Romania	23	77	66
Senegal	53	47	263
Switzerland	9	91	32
Togo	34	66	47
Tunisia	66	34	210
Turkey	60	40	575
UK	12	88	139
Vietnam	37	63	256
Asia (Other)	25	75	320

Note : The Observations are first generation immigrants (with or without children) followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first.

Table A4: Descriptive statistics - First Generation Immigrants - *Sample first* - All sample

Gender		
Women	3284	54%
Men	2765	46%
Year of Birth		
25th percentile	1957	
Median	1966	
75th percentile	1975	
Education Level		
< Primary School	1440	24%
Primary School	1565	26%
Secondary Scool	989	16%
Higher Education	1729	29%
Not Available	326	5%
Occupation Pre-Mig		
Self Employed Agricultural	132	2%
Self Employed Non-Agricultural	328	5%
High Managerial	354	6%
Supervisory Occupations	519	9%
Lower Services	724	12%
Lower Technical	817	14%
Not Available	3175	52%
Education level father		
< Primary School	2954	49%
Primary School	1006	17%
Secondary Education	465	8%
Higher Education	949	16%
Not available	675	11%
Occupation Father		
Self Employed Agricultural	774	13%
Self Employed Non-Agricultural	1075	18%
High Managerial	738	12%
Supervisory Occupations	620	10%
Lower Services	679	11%
Lower Technical	1822	30%
Not Available	341	6%

Note : The Observations are parents of second generation immigrants followed in sample first. I report the gender of the parent being interviewed, year of birth, education level and occupation before migration. I also report the occupation and education level of their father, i.e. the grandfather of second generation immigrants. The second column reports number of observations and the third the percentage of all observations they represent. The entry 2954 should be read as follows; among first generation immigrants, 2954 had a father who completed less than primary school.

Table A5: Descriptive statistics - First Generation Immigrants - *Sample Second*

	Fathers	Mothers
	Year of birth	
25th percentile	1933	1936
Median	1941	1946
75th percentile	1948	1952
	Year of arrival	
25th percentile	1961	1962
Median	1968	1970
75th percentile	1973	1976
	Education level	
$<$ Primary School	2242	2530
Primary School	735	644
Secondary Scool	163	176
Higher Education	246	182
Not available	348	202

Note : The Observations are parents of second generation immigrants followed in sample second. I report for fathers (left column) and mothers (right column), the year of birth, the year of arrival in France and the education level. The entry 182 should be read as follows; among second generation immigrants, 182 have a mother with a higher education degree.

10 Additional evidence on transition between pre-migration and first job

10.1 Various definitions of "high" and "low"

Table A6: Transition matrix - Before migration to first job - *Grandparents definition*

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	791	30	821	96.35	3.65	49.94
$S_{P,0} = H$	542	281	823	65.86	34.14	50.06
Total	1333	311	1644	81.08	18.92	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father.

Table A7: Transition matrix - Before migration to first job - *Parents definition*

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	482	41	523	92.16	7.84	59.91
$S_{P,0} = H$	240	110	350	68.57	31.43	40.09
Total	722	151	873	82.70	17.30	100

Note : The Observations are first generation male immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation.

Table A8: Transition matrix - Before migration to first job - *Grandparents definition*

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	1174	137	1311	89.55	10.45	47.41
$S_{P,0} = H$	1014	440	1454	69.74	30.26	52.59
Total	2188	577	2765	79.13	20.87	100

Note : The Observations are first generation male immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father.

Table A9: Transition matrix - Pre-Migration to current job - *Sample Second*

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,0} = L$	2829	119	2948	95.96	4.04	88.13
$S_{P,0} = H$	205	192	397	51.64	48.36	11.87
Total	3034	311	3345	90.70	9.30	100

Note : The Observations are fathers of second generation immigrants followed in sample second. High and low in period 2 refers to the education level of the father while high and low in period 2 refers to the occupation of the father.

10.2 Various origins

Table A10: Transition matrix - Before migration to first job - *Parents definition* - Europe

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	123	2	125	98.40	1.60	61.27
$S_{P,0} = H$	40	39	79	50.63	49.37	38.73
Total	163	41	204	79.90	20.10	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Europe.

Table A11: Transition matrix - Before migration to first job - *Parents definition* - Asia

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	86	3	89	96.63	3.37	62.24
$S_{P,0} = H$	42	12	54	77.78	22.22	37.76
Total	128	15	143	89.51	10.49	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Asia.

Table A12: Transition matrix - Before migration to first job - *Parents definition* - Africa

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	42	2	44	95.45	4.55	45.83
$S_{P,0} = H$	38	14	52	73.08	26.92	54.17
Total	80	16	96	83.33	16.67	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Africa.

Table A13: Transition matrix - Before migration to first job - *Parents definition* - Maghreb

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	84	3	87	96.55	3.45	73.11
$S_{P,0} = H$	26	6	32	81.25	18.75	26.89
Total	110	9	119	92.44	7.56	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Maghreb.

Table A14: Transition matrix - Before migration to first job - *Grandparents definition* - Europe

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	212	7	219	96.80	3.20	49.89
$S_{P,0} = H$	111	109	220	50.45	49.55	50.11
Total	323	116	439	73.58	26.42	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Europe.

Table A15: Transition matrix - Before migration to first job - *Grandparents definition* - Asia

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	216	5	221	97.74	2.26	54.17
$S_{P,0} = H$	154	33	187	82.35	17.65	45.83
Total	370	38	408	90.69	9.31	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Asia.

Table A16: Transition matrix - Before migration to first job - *Grandparents definition* - Africa

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	143	7	150	95.33	4.67	44.51
$S_{P,0} = H$	142	45	187	75.94	24.06	55.49
Total	285	52	337	84.57	15.43	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Africa.

Table A17: Transition matrix - Before migration to first job - *Grandparents definition* - Maghreb

	$S_{P,1}=L$	$S_{P,1}=H$	Total	$S_{P,1}=L$	$S_{P,1}=H$	Total
$S_{P,0}=L$	200	7	207	96.62	3.38	58.47
$S_{P,0}=H$	107	40	147	72.79	27.21	41.53
Total	307	47	354	86.72	13.28	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals from Maghreb.

10.3 Various waves of migration

Table A18: Transition matrix - Before migration to first job - *Parents definition* - Arrival between 1950 and 1970

	$S_{P,1}=L$	$S_{P,1}=H$	Total	$S_{P,1}=L$	$S_{P,1}=H$	Total
$S_{P,0}=L$	31	1	32	96.88	3.12	86.49
$S_{P,0}=H$	3	2	5	60.00	40.00	13.51
Total	34	3	37	91.89	8.11	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1950 and 1970.

Table A19: Transition matrix - Before migration to first job - *Parents definition* - Arrival between 1970 and 1990

	$S_{P,1}=L$	$S_{P,1}=H$	Total	$S_{P,1}=L$	$S_{P,1}=H$	Total
$S_{P,0}=L$	248	8	256	96.88	3.12	62.14
$S_{P,0}=H$	103	53	156	66.03	33.97	37.86
Total	351	61	412	85.19	14.81	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1970 and 1990.

Table A20: Transition matrix - Before migration to first job - *Parents definition* - Arrival between 1990 and 2010

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	60	1	61	98.36	1.64	41.22
$S_{P,0} = H$	51	36	87	58.62	41.38	58.78
Total	111	37	148	75.00	25.00	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the parents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1990 and 2010.

Table A21: Transition matrix - Before migration to first job - *Grandparents definition* - Arrival between 1950 and 1970

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	31	2	33	93.94	6.06	86.84
$S_{P,0} = H$	3	2	5	60.00	40.00	13.16
Total	34	4	38	89.47	10.53	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1950 and 1970.

Table A22: Transition matrix - Before migration to first job - *Grandparents definition* - Arrival between 1970 and 1990

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	346	14	360	96.11	3.89	59.70
$S_{P,0} = H$	161	82	243	66.26	33.74	40.30
Total	507	96	603	84.08	15.92	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1970 and 1990.

Table A23: Transition matrix - Before migration to first job - *Grandparents definition* - Arrival between 1990 and 2010

	$S_{P,1} = L$	$S_{P,1} = H$	Total	$S_{P,1} = L$	$S_{P,1} = H$	Total
$S_{P,0} = L$	413	14	427	96.72	3.28	42.61
$S_{P,0} = H$	378	197	575	65.74	34.26	57.39
Total	791	211	1002	78.94	21.06	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and Low in period 0 refer to the grandparents definition used for sample first. High and low in period 1 refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1990 and 2010.

11 Additional evidence on transition in France

11.1 Various definitions of "high" and "low"

Table A24: Transition matrix - From arrival to current job - *Grandparents definition*

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	1204	129	1333	90.32	9.68	81.08
$S_{P,1} = H$	33	278	311	10.61	89.39	18.92
Total	1237	407	1644	75.24	24.76	100
Natives	2325	1456	3781	61.49	38.51	100.00

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18.

Table A25: Transition matrix - From arrival to current job - *Parents definition*

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	631	91	722	87.40	12.60	82.70
$S_{P,1} = H$	19	132	151	12.58	87.42	17.30
Total	650	223	873	74.46	25.54	100
Natives	609	419	1028	59.24	40.76	100

Note : The Observations are first generation male immigrants followed in sample first. High and low in both periods refer to occupations. The observations in the last row are native men.

Table A26: Transition matrix - From arrival to current job - *Grandparents definition*

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	1951	237	2188	89.17	10.83	79.13
$S_{P,1} = H$	50	527	577	8.67	91.33	20.87
Total	2001	764	2765	72.37	27.63	100
Natives	2325	1456	3781	61.49	38.51	100.00

Note : The Observations are first generation male immigrants followed in sample first. High and low in both periods refer to occupations. The observations in the last row are native men.

11.2 Various origins

Table A27: Transition matrix - From arrival to current job - *Parents definition* - Europe

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	173	30	203	85.22	14.78	81.53
$S_{P,1} = H$	4	42	46	8.70	91.30	18.47
Total	177	72	249	71.08	28.92	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals from Europe.

Table A28: Transition matrix - From arrival to current job - *Parents definition* - Asia

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	171	17	188	90.96	9.04	85.84
$S_{P,1} = H$	7	24	31	22.58	77.42	14.16
Total	178	41	219	81.28	18.72	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals from Asia.

Table A29: Transition matrix - From arrival to current job - *Parents definition* - Africa

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	111	17	128	86.72	13.28	81.53
$S_{P,1} = H$	4	25	29	13.79	86.21	18.47
Total	115	42	157	73.25	26.75	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18. Observations are restricted to individuals from Africa.

Table A30: Transition matrix - From arrival to current job - *Parents definition* - Maghreb

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	158	21	179	88.27	11.73	90.40
$S_{P,1} = H$	3	16	19	15.79	84.21	9.60
Total	161	37	198	81.31	18.69	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18. Observations are restricted to individuals from Maghreb.

Table A31: Transition matrix - From arrival to current job - *Grandparents definition* - Europe

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	386	58	444	86.94	13.06	75.90
$S_{P,1} = H$	7	134	141	4.96	95.04	24.10
Total	393	192	585	67.18	32.82	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals from Europe.

Table A32: Transition matrix - From arrival to current job - *Grandparents definition* - asia

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	566	51	617	91.73	8.27	85.22
$S_{P,1} = H$	13	94	107	12.15	87.85	14.78
Total	579	145	724	79.97	20.03	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals from Asia.

Table A33: Transition matrix - From arrival to current job - *Grandparents definition* - Africa

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	466	55	521	89.44	10.56	80.15
$S_{P,1} = H$	14	115	129	10.85	89.15	19.85
Total	480	170	650	73.85	26.15	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals from Africa.

Table A34: Transition matrix - From arrival to current job - *Grandparents definition* - Maghreb

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	462	60	522	88.51	11.49	82.46
$S_{P,1} = H$	11	100	111	9.91	90.09	17.54
Total	473	160	633	74.72	25.28	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18. Observations are restricted to individuals from Maghreb.

11.3 Various waves of migration

Table A35: Transition matrix - From arrival to current job - *Parents definition* - Arrival between 1950 and 1970

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	59	18	77	76.62	23.38	89.53
$S_{P,1} = H$	1	8	9	11.11	88.89	10.47
Total	60	26	86	69.77	30.23	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18. Observations are restricted to individuals who arrived in France between 1950 and 1970.

Table A36: Transition matrix - From arrival to current job - *Parents definition* - Arrival between 1970 and 1990

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	464	56	520	89.23	10.77	83.60
$S_{P,1} = H$	14	88	102	13.73	86.27	16.40
Total	478	144	622	76.85	23.15	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1970 and 1990.

Table A37: Transition matrix - From arrival to current job - *Parents definition* - Arrival between 1990 and 2010

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	107	17	124	86.29	13.71	76.07
$S_{P,1} = H$	4	35	39	10.26	89.74	23.93
Total	111	52	163	68.10	31.90	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1990 and 2010.

Table A38: Transition matrix - From arrival to current job - *Grandparents definition* - Arrival between 1950 and 1970

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	70	21	91	76.92	23.08	85.85
$S_{P,1} = H$	3	12	15	20.00	80.00	14.15
Total	73	33	106	68.87	31.13	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1950 and 1970.

Table A39: Transition matrix - From arrival to current job - *Grandparents definition* - Arrival between 1970 and 1990

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	832	126	958	86.85	13.15	78.46
$S_{P,1} = H$	28	235	263	10.65	89.35	21.54
Total	860	361	1221	70.43	29.57	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The observations in the last row are native men who have at least one child who is at least 18. Observations are restricted to individuals who arrived in France between 1970 and 1990.

Table A40: Transition matrix - From arrival to current job - *Grandparents definition* - Arrival between 1990 and 2010

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	1047	90	1137	92.08	7.92	79.23
$S_{P,1} = H$	19	279	298	6.38	93.62	20.77
Total	1066	369	1435	74.29	25.71	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. Observations are restricted to individuals who arrived in France between 1990 and 2010.

11.4 Those who were previously high

Table A41: Transition matrix - From arrival to current job - *Grandparents definition* - Previously High Status

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	445	97	542	82.10	17.90	65.86
$S_{P,1} = H$	26	255	281	9.25	90.75	34.14
Total	471	352	823	57.23	42.77	100

Note : The Observations are fathers of second generation immigrants followed in sample first. High and low in both periods refer to the occupation of the father. The sample is limited to fathers with a high status in the origin country.

Table A42: Transition matrix - From arrival to current job - *Parents definition* - Previously High Status

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	195	45	240	81.25	18.75	68.57
$S_{P,1} = H$	7	103	110	6.36	93.64	31.43
Total	202	148	350	57.71	42.29	100

Note : The Observations are first generation male immigrants followed in sample first. High and low in both periods refer to occupations. The sample is limited to males with a high status in the origin country.

Table A43: Transition matrix - From arrival to first job - *Grandparents definition* - Previously High Status

	$S_{P,2} = L$	$S_{P,2} = H$	Total	$S_{P,2} = L$	$S_{P,2} = H$	Total
$S_{P,1} = L$	866	148	1014	85.40	14.60	69.74
$S_{P,1} = H$	30	410	440	6.82	93.18	30.26
Total	896	558	1454	61.62	38.38	100

Note : The Observations are first generation male immigrants followed in sample first. High and low in both periods refer to occupations. The sample is limited to males with a high status in the origin country.

12 Additional evidence on the resurgence of pre-migration background

12.1 From sample first

Using parents definition

Table A44: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.124***	0.130***	0.143***		
	(0.033)	(0.044)	(0.036)		
$S_{P,1} = H$		0.075*			
		(0.055)			
$S_{P,2} = H$	0.136***				
	(0.033)				
Houhesold $			0.002**		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				-0.041	
				(0.096)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.117***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.227***	
				(0.071)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.007
					(0.066)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.091***
					(0.036)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.283***
					(0.039)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin and number of siblings. Standard errors are clustered at the household level.

Table A45: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.119***	0.132***	0.138***		
	(0.033)	(0.043)	(0.037)		
$S_{P,1} = H$		0.056			
		(0.055)			
$S_{P,2} = H$	0.132***				
	(0.033)				
Houhesold $			0.002**		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.001	
				(0.098)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.125***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.198***	
				(0.070)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.021
					(0.065)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.091***
					(0.035)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.272***
					(0.040)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies) for the parent being interviewed. Standard errors are clustered at the household level.

Table A46: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.120***	0.143***	0.132***		
	(0.033)	(0.044)	(0.037)		
$S_{P,1} = H$		0.054			
		(0.057)			
$S_{P,2} = H$	0.123***				
	(0.033)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.001	
				(0.113)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.137***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.206***	
				(0.075)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.006
					(0.065)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.091***
					(0.035)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.266***
					(0.040)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies) and year of moving to France for the parent being interviewed. Standard errors are clustered at the household level.

Table A47: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.188***	0.216***	0.165***		
	(0.045)	(0.055)	(0.051)		
$S_{P,1} = H$		0.046			
		(0.085)			
$S_{P,2} = H$	0.090***				
	(0.047)				
Houhesold $			0.004***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				-0.205	
				(0.182)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.196***	
				(0.057)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.312***	
				(0.089)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					-0.036
					(0.076)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.148***
					(0.048)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.319***
					(0.055)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control mean educ (at moving)	Yes	Yes	Yes	Yes	Yes
Mean	0.561	0.535	0.555	0.535	0.561
N	1549	861	1155	861	1549
N High	473	234	535	234	473

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France of the parent being interviewed and mean level of education in the country of origin the year of moving to France of the parent being interviewed. Standard errors are clustered at the household level.

Table A48: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.182***	0.214***	0.144***		
	(0.048)	(0.063)	(0.057)		
$S_{P,1} = H$		0.096			
		(0.101)			
$S_{P,2} = H$	0.094***				
	(0.051)				
Houhesold $			0.004**		
			(0.002)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				-0.317**	
				(0.163)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.174***	
				(0.064)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.396***	
				(0.105)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					-0.011
					(0.091)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.151***
					(0.051)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.303***
					(0.062)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control mean educ (at birth)	Yes	Yes	Yes	Yes	Yes
Mean	0.600	0.583	0.588	0.583	0.600
N	1181	665	898	665	1181
N High	417	208	314	208	417

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France for the parent being interviewed and mean level of education in the country of origin the year of birth of the parent being interviewed. Standard errors are clustered at the household level.

Table A49: Resurgence - *Parents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.120***	0.151***	0.125***		
	(0.033)	(0.044)	(0.037)		
$S_{P,1} = H$		0.033			
		(0.059)			
$S_{P,2} = H$	0.114***				
	(0.034)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				-0.012	
				(0.118)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.145***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.193***	
				(0.074)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					-0.004
					(0.065)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.091***
					(0.035)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.257***
					(0.040)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control sensitive area	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France for the parent being interviewed and a dummy for living in a Urban Sensitive Area (ZUS). Standard errors are clustered at the household level.

Using grandparents definition

Table A50: Resurgence - *Grandparents Definition* - Baseline

Obtaining the Baccalaureat					
$S_{P,0} = H$	0.134***	0.174***	0.132***		
	(0.021)	(0.032)	(0.025)		
$S_{P,1} = H$		0.121***			
		(0.041)			
$S_{P,2} = H$	0.143***				
	(0.025)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.101	
				(0.078)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.169***	
				(0.035)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.303***	
				(0.039)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.142***
					(0.04)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.134***
					(0.024)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.277***
					(0.029)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Mean	0.575	0.584	0.563	0.584	0.575
N	4278	1694	3129	1694	4278
N High	1442	561	1083	561	1442

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender and dummies for country of origin. Standard errors are clustered at the household level.

Table A51: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.118***	0.154***	0.119***		
	(0.022)	(0.033)	(0.026)		
$S_{P,1} = H$		0.101***			
		(0.042)			
$S_{P,2} = H$	0.136***				
	(0.025)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.082	
				(0.078)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.150***	
				(0.036)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.262***	
				(0.042)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.136***
					(0.040)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.118***
					(0.024)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.254***
					(0.030)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Mean	0.575	0.584	0.563	0.584	0.575
N	4278	1694	3129	1694	4278
N High	1442	561	1083	561	1442

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin and number of siblings. Standard errors are clustered at the household level.

Table A52: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.118***	0.148***	0.117***		
	(0.022)	(0.034)	(0.026)		
$S_{P,1} = H$		0.091**			
		(0.042)			
$S_{P,2} = H$	0.123***				
	(0.025)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.086	
				(0.076)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.147***	
				(0.037)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.240***	
				(0.044)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.125***
					(0.040)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.118***
					(0.024)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.240***
					(0.030)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Mean	0.575	0.584	0.563	0.584	0.575
N	4278	1694	3129	1694	4278
N High	1442	561	1023	561	1442

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies) for the parent being interviewed. Standard errors are clustered at the household level.

Table A53: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.120***	0.143***	0.132***		
	(0.033)	(0.044)	(0.037)		
$S_{P,1} = H$		0.054			
		(0.057)			
$S_{P,2} = H$	0.123***				
	(0.033)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.001	
				(0.113)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.137***	
				(0.046)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.206***	
				(0.075)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.006
					(0.065)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.091***
					(0.035)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.266***
					(0.040)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Mean	0.613	0.589	0.607	0.589	0.613
N	2142	1120	1570	1120	2142
N High	832	400	580	400	832

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies) and year of moving to France for the parent being interviewed. Standard errors are clustered at the household level.

Table A54: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.148***	0.197***	0.147***		
	(0.028)	(0.042)	(0.033)		
$S_{P,1} = H$		0.112***			
		(0.060)			
$S_{P,2} = H$	0.114***				
	(0.034)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.064	
				(0.110)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.188***	
				(0.045)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.331***	
				(0.062)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.119***
					(0.049)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.150***
					(0.030)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.258***
					(0.043)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control mean educ (at moving)	Yes	Yes	Yes	Yes	Yes
Mean	0.536	0.543	0.521	0.543	0.536
N	3297	1299	2428	1299	3297
N High	899	347	648	347	899

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France of the parent being interviewed and mean level of education in the country of origin the year of moving to France of the parent being interviewed. Standard errors are clustered at the household level.

Table A55: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.184***	0.219***	0.191***		
	(0.032)	(0.048)	(0.037)		
$S_{P,1} = H$		0.119**			
		(0.065)			
$S_{P,2} = H$	0.094***				
	(0.038)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.023	
				(0.132)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.202***	
				(0.051)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.376***	
				(0.064)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.017
					(0.062)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.161***
					(0.034)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.317***
					(0.042)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control mean educ (at birth)	Yes	Yes	Yes	Yes	Yes
Mean	0.568	0.570	0.553	0.570	0.568
N	2271	982	1748	982	2271
N High	740	307	543	740	307

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France for the parent being interviewed and mean level of education in the country of origin the year of birth of the parent being interviewed. Standard errors are clustered at the household level.

Table A56: Resurgence - *Grandparents Definition* - Robustness

Obtaining the Baccalaureat

$S_{P,0} = H$	0.114***	0.141***	0.110***		
	(0.023)	(0.036)	(0.027)		
$S_{P,1} = H$		0.073*			
		(0.045)			
$S_{P,2} = H$	0.113***				
	(0.026)				
Houhesold $			0.003***		
			(0.001)		
$S_{P,0} = L \,\&\, S_{P,1} = H$				0.075	
				(0.083)	
$S_{P,0} = H \,\&\, S_{P,1} = L$				0.141***	
				(0.039)	
$S_{P,0} = H \,\&\, S_{P,1} = H$				0.214***	
				(0.047)	
$S_{P,0} = L \,\&\, S_{P,2} = H$					0.115***
					(0.042)
$S_{P,0} = H \,\&\, S_{P,2} = L$					0.115***
					(0.025)
$S_{P,0} = H \,\&\, S_{P,2} = H$					0.226***
					(0.031)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Gender	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control sensitive area	Yes	Yes	Yes	Yes	Yes
Mean	0.575	0.584	0.564	0.584	0.575
N	4276	1693	3128	4276	1693
N High	1442	561	1023	1442	561

Note : The first three columns are the results of the estimation of equation 1. The first column uses binary information on father occupation at the time of the interview, the second uses information on occupation at the time of arrival and the third on monetary resources available at the household level (in hundreds of euros, this information is only available for those who are still working). The fourth and fifth columns report the results of estimating equation 2. I report the mean of the dependent variable (having obtained the baccalaureat), together with the number of observations and number of $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings, year of birth (dummies), year of moving to France for the parent being interviewed and a dummy for living in a Urban Sensitive Area (ZUS). Standard errors are clustered at the household level.

12.2 From sample second

Table A57: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.06***	0.14***	0.24***	0.13	0.21***
	(0.02)	(0.03)	(0.05)	(0.12)	(0.08)
$S_{P,2} = H$	0.02	0.11***	0.10**	0.16**	0.19***
	(0.02)	(0.03)	(0.05)	(0.07)	(0.08)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Mean	0.87	0.53	0.32	7.40	0.37
N	3335	3335	1930	666	858
N High	397	397	143	46	63
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.03	0.14***	0.13**	0.24**	0.26***
	(0.03)	(0.04)	(0.06)	(0.07)	(0.09)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.07***	0.16***	0.26***	0.25*	0.30***
	(0.02)	(0.04)	(0.06)	(0.16)	(0.10)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.08***	0.22***	0.31***	0.16*	0.30***
	(0.01)	(0.03)	(0.06)	(0.11)	(0.10)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Mean	0.87	0.53	0.32	7.40	0.37
N	3335	3335	1930	666	858
N High	397	397	143	46	63

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin and number of siblings. Standard errors are robust to general forms of heteroskedasticity.

Table A58: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.04**	0.16***	0.22***	0.22*	0.28***
	(0.02)	(0.04)	(0.07)	(0.15)	(0.10)
$S_{P,2} = H$	0.03	0.11***	0.09*	0.14**	0.29***
	(0.03)	(0.04)	(0.06)	(0.08)	(0.09)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.50	0.31	7.44	0.39
N	1824	1824	1523	511	637
N High	175	175	112	34	46
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.05	0.15***	0.10*	0.24***	0.34***
	(0.04)	(0.06)	(0.07)	(0.08)	(0.09)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.05**	0.19***	0.23***	0.47**	0.42***
	(0.03)	(0.05)	(0.08)	(0.22)	(0.13)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.05***	0.23***	0.31***	0.15	0.45***
	(0.02)	(0.05)	(0.07)	(0.12)	(0.13)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.50	0.31	7.44	0.39
N	1824	1824	1523	511	637
N High	175	175	112	34	46

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies) of the father. Standard errors are robust to general forms of heteroskedasticity.

Table A59: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.03*	0.16***	0.26***	0.19*	0.28***
	(0.02)	(0.05)	(0.07)	(0.13)	(0.11)
$S_{P,2} = H$	0.03	0.11**	0.06	0.15**	0.29***
	(0.03)	(0.05)	(0.06)	(0.09)	(0.09)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.52	0.32	7.46	0.41
N	1631	1631	1385	472	585
N High	157	157	105	31	41
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.04	0.15**	0.08	0.23**	0.32***
	(0.05)	(0.07)	(0.08)	(0.10)	(0.10)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.05**	0.188***	0.28***	0.45**	0.39***
	(0.03)	(0.07)	(0.10)	(0.25)	(0.16)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.08**	0.29***	0.37***	0.04	0.35***
	(0.04)	(0.07)	(0.11)	(0.21)	(0.19)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Mean	0.86	0.52	0.32	7.46	0.41
N	1631	1631	1385	472	585
N High	157	157	105	31	41

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies) and year of moving to France of the father. Standard errors are robust to general forms of heteroskedasticity.

Table A60: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.05**	0.17***	0.29***	0.15	0.23**
	(0.03)	(0.06)	(0.08)	(0.18)	(0.13)
$S_{P,2} = H$	0.04	0.14***	0.08	0.15*	0.27***
	(0.04)	(0.06)	(0.07)	(0.10)	(0.10)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Control mean education (at moving)	Yes	Yes	Yes	Yes	Yes
Mean	0.85	0.49	0.29	7.46	0.40
N	1464	1464	1259	423	524
N High	99	99	66	19	24
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.04	0.15**	0.08	0.23**	0.32***
	(0.05)	(0.07)	(0.08)	(0.10)	(0.10)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.05**	0.188***	0.28***	0.45**	0.39***
	(0.03)	(0.07)	(0.10)	(0.25)	(0.16)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.08**	0.29***	0.37***	0.04	0.35***
	(0.04)	(0.07)	(0.11)	(0.21)	(0.19)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Control mean education (at moving)	Yes	Yes	Yes	Yes	Yes
Mean	0.85	0.49	0.29	7.46	0.40
N	1464	1464	1259	423	524
N High	99	99	66	19	24

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies), year of moving to France of the father and mean level of education in the country of origin the year of moving to France. Standard errors are robust to general forms of heteroskedasticity.

Table A61: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.05**	0.14***	0.24***	0.14	0.24**
	(0.03)	(0.06)	(0.09)	(0.17)	(0.13)
$S_{P,2} = H$	0.03	0.14***	0.05	0.13	0.28***
	(0.04)	(0.05)	(0.07)	(0.11)	(0.10)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Control mean education (at birth)	Yes	Yes	Yes	Yes	Yes
Mean	0.85	0.49	0.29	7.46	0.41
N	1461	1461	1257	424	528
N High	99	99	65	21	28
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.04	0.15**	0.06	0.20**	0.32***
	(0.05)	(0.07)	(0.08)	(0.10)	(0.10)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.05**	0.16***	0.25***	0.43**	0.37***
	(0.03)	(0.07)	(0.10)	(0.25)	(0.17)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.07**	0.26***	0.28***	0.03	0.41***
	(0.04)	(0.07)	(0.11)	(0.19)	(0.17)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year (father)	Yes	Yes	Yes	Yes	Yes
Control moving year (father)	Yes	Yes	Yes	Yes	Yes
Control mean education (at birth)	Yes	Yes	Yes	Yes	Yes
Mean	0.85	0.49	0.29	7.46	0.41
N	1461	1461	1257	424	528
N High	99	99	65	21	28

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies), year of moving to France of the father and mean level of education in the country of origin the year of birth of the father. Standard errors are robust to general forms of heteroskedasticity.

Table A62: Resurgence - *Sample Second* - Robustness

	D-O	BAC	H-E	ln Wage	Occupation
$S_{P,0} = H$	0.03	0.16***	0.25***	0.18*	0.25**
	(0.03)	(0.05)	(0.07)	(0.14)	(0.12)
$S_{P,2} = H$	0.02	0.12***	0.06	0.16**	0.30***
	(0.03)	(0.05)	(0.06)	(0.09)	(0.09)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control sensitive area	Yes	Yes	Yes	Yes	Yes
Mean	0.87	0.52	0.32	7.46	0.41
N	1582	1582	1343	458	569
N High	145	145	97	28	38
$S_{P,0} = L \,\&\, S_{P,2} = H$	0.04	0.14***	0.06	0.23**	0.35***
	(0.04)	(0.06)	(0.07)	(0.09)	(0.09)
$S_{P,0} = H \,\&\, S_{P,2} = L$	0.04*	0.18***	0.26***	0.46**	0.39***
	(0.03)	(0.05)	(0.08)	(0.20)	(0.16)
$S_{P,0} = H \,\&\, S_{P,2} = H$	0.04*	0.26***	0.30***	0.11	0.42***
	(0.03)	(0.06)	(0.09)	(0.17)	(0.16)
Control Age	Yes	Yes	Yes	Yes	Yes
Control Age Squared	No	No	No	Yes	Yes
Control Gender	Yes	Yes	Yes	No	No
Country FE	Yes	Yes	Yes	Yes	Yes
Nb of siblings	Yes	Yes	Yes	Yes	Yes
Control birth year	Yes	Yes	Yes	Yes	Yes
Control moving year	Yes	Yes	Yes	Yes	Yes
Control sensitive area	Yes	Yes	Yes	Yes	Yes
Mean	0.87	0.52	0.32	7.46	0.41
N	1582	1582	1343	458	569
N High	145	145	97	28	38

Note : The upper part of the table summarizes the results of the estimation of equation 1. The lower part of the table reports the results of estimating equation 2. I report the mean of the dependent variable (respectively not dropping out "D-O", obtaining the baccalaureat "BAC", obtaining a higher education degree "H-E", log of wage "Ln Wage" and a binary indicator for high occupation "Occupation"), together with the total number of observations and number of observations with $S_{P,0}$= H. All the estimations control for age, gender, dummies for country of origin, number of siblings and year of birth (dummies), year of moving to France of the father and a dummy for living in an ethnic enclave. Standard errors are robust to general forms of heteroskedasticity.

13 Additional evidence on channels

Table A63: Returns to experience and Language classes - Additional controls

	Immigrants with children			All Immigrants		
	Immigrants		Natives	Immigrants		Natives
Experience	-0.161	-0.063	0.121***	0.037	0.057	0.045***
	(0.015)	(0.460)	(0.426)	(0.254)	(0.292)	(0.004)
Experience sq	0.003	0.001	-0.002**	-0.001	-0.001	-0.001***
	(0.000)	(0.008)	(0.007)	(0.004)	(0.005)	(0.000)
Nb observations	506	506	246	1747	1747	1213
Control Education	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	No	Yes		No	Yes	
Control birth year (father)	Yes	Yes		Yes	Yes	
Language classes	0.023	0.038		-0.025	-0.002	
	(0.051)	(0.052)		(0.029)	(0.029)	
Control Education	Yes	Yes		Yes	Yes	
Control Age	Yes	Yes		Yes	Yes	
Country FE	No	Yes		No	Yes	
Control birth year (father)	Yes	Yes		Yes	Yes	
Nb observations	356	356		1158	1158	
N High *Parents definition*	87	87		360	360	
N High *Gdparents definition*	140	140		568	568	

Note: The first two columns of the upper part of the table report the results of the regression of equation 8. The coefficients reported refer to experience in France (calculated as the number of years since arrival). The sample comprises fathers of second generation immigrants. Controls include dummies for level of education and birth year of the father. Certain specifications include dummies for country of origin. The third column estimates a similar specification for natives with adult children where potential experience is calculated as age minus 18. The last three columns of the upper part of the table replicates these results on a sample of all male immigrants and natives. The first two columns of the lower part of the table reports the estimation of equation 9 on a sample of fathers of second generation immigrants who were not fluent in French at the time of arrival. All specifications include dummies for education level and birth year of the father; some include country of origin fixed effects. The last two rows specifiy the number of $S_{P,0} = H$ in the sample according to both the parents and grandparents definition. The last three columns of the lower part of the table replicates these results on a sample of all male immigrants.